



# Reconstruction des vitesses propres des galaxies: méthodes et applications aux observations.

Guilhem Lavaux

## ► To cite this version:

Guilhem Lavaux. Reconstruction des vitesses propres des galaxies: méthodes et applications aux observations.. Cosmologie et astrophysique extra-galactique [astro-ph.CO]. Université Paris Sud - Paris XI, 2008. Français. NNT: . tel-00412146

**HAL Id: tel-00412146**

**<https://theses.hal.science/tel-00412146>**

Submitted on 1 Sep 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NUMERO D'ORDRE :

THÈSE DE DOCTORAT EN SCIENCE UNIVERSITÉ DE PARIS XI  
Faculté des sciences d'Orsay, spécialité : astrophysique

Ecole doctorale d'astronomie et d'astrophysique d'Île-de-France

présentée par  
Guilhem LAVAUX

soutenue le 2 Juillet 2008

pour l'obtention du grade de  
DOCTEUR DE L'UNIVERSITÉ PARIS XI

<p><b>Reconstruction des vitesses propres des galaxies : méthodes et applications aux observations</b></p>
--

---

Composition du jury de thèse (ordre alphabétique) :

M.	Francis	BERNARDEAU,	CEA SAP Saclay,	Directeur
Pr.	Alain	BLANCHARD,	Université Paul Sabatier,	Rapporteur
M.	Stéphane	COLOMBI,	IAP & CNRS,	Examineur
Dr.	Roya	MOHAYAEI,	IAP & CNRS,	Invité
M.	Jean-Loup	PUGET,	IAS & Univ. Paris 11,	Examineur
Pr.	Joseph	SILK,	Univ. of Oxford	Invité
M.	Eric	THIÉBAUT,	CRAL & Univ. de Lyon I,	Examineur
Pr.	Benjamin	WANDEL,	University of Illinois,	Rapporteur





# Résumé

Bien que nous ayons accès à des catalogues de galaxies très détaillés, notre compréhension de la distribution spatiale de la matière noire reste encore limitée. D'importantes informations à son sujet sont cachées dans les vitesses propres des galaxies, qui reflètent la dynamique de la matière noire à différentes échelles. Malheureusement, ces vitesses sont très difficiles à observer. Nous présentons ici une approche différente pour “mesurer” ces vitesses par l'intermédiaire de méthodes de reconstruction des champs de vitesse. Nous utilisons en particulier la reconstruction dite de Monge-Ampère-Kantorovitch (MAK). Nous testons cette méthode sur des simulations à  $N$ -corps ainsi que sur des catalogues virtuels de galaxies. Nous vérifions sa fiabilité par la comparaison des vitesses reconstruites aux vitesses simulées et aussi à travers la mesure de la densité moyenne de matière de ces univers.

Après avoir testé cette méthode, nous l'utilisons sur un vrai catalogue de galaxies : le 2MASS Redshift survey. Après l'avoir corrigé des effets observationnels connus, nous étudions l'origine de la vitesse du Groupe Local par rapport au fond diffus cosmologique. Nous montrons que plus de la moitié de notre vitesse est due à des structures situées à plus de  $40h^{-1}\text{Mpc}$ . Une fois étudié le mouvement d'ensemble des structures locales, nous comparons directement les vitesses reconstruites et les distances observées dans notre voisinage de  $30h^{-1}\text{Mpc}$ . Nous proposons une estimation indépendante du paramètre de densité. Cette estimation peut être utilisée afin de réduire les dégénérescences dans l'espace des paramètres du modèle d'univers à base de matière noire froide.

# Abstract

Although we now have access to very detailed galaxy catalogs, our knowledge of the spatial distribution of dark matter remains limited. A valuable source of information lies in the peculiar velocity field of the galaxies, which traces the dynamics of the dark matter at various scales. Unfortunately, these velocities are very difficult to observe. We present here an alternative approach to measure these velocities which relies on reconstructing them directly from redshift catalogs. In particular, we use the so-called Monge-Ampère-Kantorovitch reconstruction method. We test this method on  $N$ -body simulations and mock galaxy catalogs. We check its reliability by comparing the reconstructed and simulated peculiar velocities and also the mean matter density.

Having tested the method, we apply it to a real galaxy catalog : the 2MASS Redshift survey. Once known observational effects have been accounted for, we study the origin of the Cosmic Microwave Background (CMB) dipole, precisely the motion of the Local Group relative to it. We demonstrate that more than half of our dipole motion is due to structures which lie beyond  $40h^{-1}\text{Mpc}$ . Having studied the bulk flow, we make a direct comparison of reconstructed velocities and the observed distances in our neighborhood of  $30h^{-1}\text{Mpc}$  and provide an independent estimation of the density parameter. This can be used to reduce the degeneracy in the parameter space of the present cosmological cold dark matter model.



# Remerciements

Beaucoup de gens ont contribué de diverses manières au succès de cette thèse.

En premier lieu Roya et Stéphane qui m'ont encadré pendant ces trois ans. Ils m'ont constamment soutenu et encouragé afin que je progresse dans le métier de chercheur. Je remercie Brent Tully pour sa grande connaissance qu'il a pu partager avec moi de l'Univers Local. Je le remercie aussi de m'avoir invité chaleureusement à l'Institute for Astronomy à Hawaii. C'est un endroit formidable. Cette thèse n'aurait jamais pu commencer sans l'aide de Joe Silk, que j'avais eu la chance de rencontrer en 2003. Au titre de l'IAP, je souhaiterais remercier le précieux concours de Christophe Alard, Simon Prunet et Christophe Pichon.

Un grand merci à Michał Chodorowski pour de nombreuses discussions passionnantes et pour m'avoir invité à passer quelques jours à Varsovie au Copernicus Center. Je remercie aussi les gens du Copernicus Center (en particulier Radek & Maciek) pour m'avoir accompagné pendant ces deux semaines à Varsovie.

Je remercie John Huchra et les membres du 2MASS Redshift Survey pour m'avoir communiqué leurs précieuses données. Les membres du projet HORIZON et OTARIE pour le soutien financier et technique qui a été nécessaire durant ces trois ans. Magique 2 (et son soutien humain Stéphane Rouberol :) pour m'avoir bien servi pour faire exécuter les innombrables calculs de reconstruction que j'ai effectué pendant ma thèse.

Merci enfin à ma famille pour leur soutien moral durant ces trois ans. Pour finir, merci à tous mes amis et à Elisabetta (je t'embrasse très fort).



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Notre Univers : cosmographie et dynamique locale</b>	<b>3</b>
2.1	Notre Univers Local / catalogues de galaxie . . . . .	3
2.2	L'Univers homogène . . . . .	7
2.2.1	Métrie de Friedmann-Lemaître-Robertson-Walker (FLRW) . . . . .	7
2.2.2	Formes réduites – contenu de l'Univers . . . . .	10
2.2.3	Quelques modèles d'Univers usuels . . . . .	10
2.2.4	Distances dans l'Univers . . . . .	11
2.2.5	Mesures les plus récentes des paramètres cosmologiques . . . . .	13
2.2.6	Cinématique des objets dans un Univers de FLRW . . . . .	13
2.3	L'Univers inhomogène : dynamique linéaire des fluctuations . . . . .	14
2.3.1	L'approche Eulérienne . . . . .	14
2.3.2	Théorie linéaire . . . . .	15
2.3.3	L'approche Lagrangienne : l'approximation de Zel'dovich . . . . .	17
2.4	Intérêt et obtention des champs de vitesse . . . . .	18
2.4.1	Intérêt des champs de vitesse . . . . .	18
2.4.2	Les distances dans l'Univers / Mesurer les vitesses . . . . .	19
2.4.3	Les méthodes de reconstruction . . . . .	21
2.4.4	La méthode POTENT . . . . .	22
2.4.5	La méthode de moindre action . . . . .	22
2.4.6	La reconstruction de Monge-Ampère-Kantorovitch (MAK) . . . . .	26
<b>3</b>	<b>Test de la reconstruction MAK sur une simulation</b>	<b>29</b>
3.1	La simulation $N$ corps . . . . .	29
3.1.1	Les paramètres utilisés . . . . .	29
3.1.2	Fluctuation à grande échelle du champ de vitesse . . . . .	30
3.2	Identification des halos . . . . .	31
3.3	Reconstruction du champ de déplacement . . . . .	31
3.3.1	De la simulation . . . . .	31
3.3.2	Sur un catalogue de halo . . . . .	33
3.4	Reconstruction du champ de vitesse . . . . .	35
3.4.1	Champ de vitesse de la simulation . . . . .	35
3.4.2	Champ de vitesse sur FullMock . . . . .	37
3.4.3	Statistique sur la direction . . . . .	37
3.5	Reconstruction des fluctuations primordiales . . . . .	40
3.5.1	Retrouver les conditions initiales de la simulation avec MAK . . . . .	40
3.5.2	Le modèle du collapse sphérique . . . . .	42
3.5.3	Application au champ de déplacement MAK . . . . .	46
3.6	Conclusion . . . . .	47
<b>4</b>	<b>Analyse bayésienne des champs de vitesse</b>	<b>51</b>
4.1	Les erreurs dans la reconstruction MAK . . . . .	51
4.1.1	Distribution des erreurs . . . . .	51

4.1.2	Outils pour quantifier les erreurs . . . . .	52
4.2	Mesurer $\Omega_m$ . . . . .	55
4.2.1	La méthode des moments . . . . .	55
4.2.2	La méthode de la fonction de vraisemblance . . . . .	55
4.3	Analyse bayésienne . . . . .	56
4.3.1	Le modèle bayésien . . . . .	56
4.3.2	Méthode numérique de calcul de la fonction de vraisemblance . . . . .	61
4.3.3	Test de la méthode sur des données générées aléatoirement . . . . .	62
4.3.4	Analyse bayésienne : introduction de la corrélation . . . . .	63
<b>5</b>	<b>Etude des biais observationnels avec des catalogues virtuels de galaxies</b>	<b>67</b>
5.1	Catalogues virtuels de base pour l'étude . . . . .	68
5.2	La relation masse-luminosité . . . . .	69
5.2.1	Absence de traceurs de masse . . . . .	70
5.2.2	Rapport masse-luminosité de groupes de galaxies . . . . .	72
5.2.3	Limitation en magnitude . . . . .	76
5.3	Distortion de décalage vers le rouge (redshift) . . . . .	79
5.4	Effets dues à la géométrie du catalogue . . . . .	86
5.4.1	Le plan galactique de la voie lactée . . . . .	86
5.4.2	Le domaine lagrangien . . . . .	87
5.4.3	La variance cosmique . . . . .	93
5.5	Erreurs sur les mesures de distance . . . . .	98
5.5.1	A-t-on réellement besoin d'une analyse de vraisemblance ? . . . . .	98
5.5.2	Résultat . . . . .	98
5.6	Conclusion sur les systématiques . . . . .	101
<b>6</b>	<b>Le champs de vitesse du 2MASS Redshift survey (2MRS)</b>	<b>103</b>
6.1	Le catalogue de distance NBG-3k . . . . .	103
6.1.1	Présentation du catalogue . . . . .	103
6.1.2	Traitement du catalogue en vue de la reconstruction . . . . .	104
6.1.3	Le champs de vitesse reconstruit . . . . .	106
6.2	Le catalogue de galaxies avec décalage vers le rouge : 2MASS Redshift Survey . . . . .	107
6.2.1	Présentation du catalogue . . . . .	107
6.2.2	Distributions des galaxies du catalogue . . . . .	110
6.2.3	Préparation du catalogue : phase I . . . . .	110
6.2.4	Préparation du catalogue : phase II . . . . .	119
6.2.5	Comparaison du champ de densité du NBG-3k et du 2MASS . . . . .	124
6.2.6	Champ de vitesse reconstruit . . . . .	124
6.3	La vitesse d'ensemble locale et l'origine du dipôle du CMB . . . . .	125
6.3.1	Mesurer la vitesse d'ensemble : le méthode $\chi^2$ . . . . .	125
6.3.2	Analyse de la vitesse d'ensemble dans une simulation $\Lambda$ CDM . . . . .	130
6.3.3	La vitesse d'ensemble observée par le catalogue NBG-3k . . . . .	133
6.3.4	La vitesse d'ensemble reconstruite à partir du catalogue 2MRS . . . . .	134
6.4	Comparaison aux vitesses observées et paramètres cosmologiques . . . . .	136
6.4.1	Analyse des vitesses simulées et reconstruites d'une simulation $\Lambda$ CDM . . . . .	145
6.4.2	Vitesse observées et reconstruites dans le catalogue 2MRS/NBG-3k . . . . .	145
6.5	Conclusion . . . . .	147
<b>7</b>	<b>Conclusion et perspectives</b>	<b>149</b>
<b>A</b>	<b>Notation utilisée dans le manuscrit de thèse</b>	<b>155</b>
A.1	Vecteurs - scalaire . . . . .	155

A.2	Transformée de Fourier . . . . .	155
A.3	Projection Aitoff . . . . .	155
A.4	Fonction de Bessel sphérique . . . . .	156
<b>B</b>	<b>Le problème de minimisation</b>	<b>157</b>
B.1	Un algorithme stochastique . . . . .	157
B.1.1	Approche dense . . . . .	157
B.1.2	Version sparse . . . . .	157
B.2	L'algorithme Auction . . . . .	158
B.2.1	La machine de Hénon et l'idée derrière l'algorithme Auction . . . . .	158
B.2.2	Théorie . . . . .	158
B.2.3	Amélioration . . . . .	160
B.2.4	Implémentation parallèle . . . . .	160
B.2.5	Version "sparse"–"dense" . . . . .	161
B.2.6	Test de performance . . . . .	161
<b>C</b>	<b>Fonction de corrélations</b>	<b>163</b>
C.1	Champ de densité . . . . .	163
C.2	Fonction de corrélation du champ de vitesse lissé . . . . .	165
C.3	Variance de l'estimateur radial du bulk-flow . . . . .	166
<b>D</b>	<b>Fonction de masse</b>	<b>171</b>
D.1	Présentation de la théorie de Press-Schechter (forme BBKS) . . . . .	171
D.2	La modification de Sheth, Mo & Tormen . . . . .	171
D.3	Compter la fraction manquante . . . . .	172
<b>E</b>	<b>Filtrage adaptatif</b>	<b>175</b>
E.1	Le lissage adaptatif général . . . . .	175
E.2	Le lissage avec conservation de l'intégrale du champ . . . . .	176
E.2.1	Formalisme général . . . . .	176
E.2.2	Application à l'interpolation du champ de vitesse et de la masse . . . . .	176
<b>F</b>	<b>Proceedings</b>	<b>179</b>
<b>G</b>	<b>Proceedings référés</b>	<b>189</b>
<b>H</b>	<b>Article référé</b>	<b>195</b>
<b>I</b>	<b>Article soumis</b>	<b>225</b>





# 1 Introduction

L'ASTRONOMIE, en tant qu'étude des lois des astres, est considérée comme la plus ancienne des sciences. On dispose de vestiges préhistoriques montrant que les hommes se sont intéressés très tôt à l'origine et au mouvement des "étoiles" dans le ciel. Cependant, ce n'est qu'à l'époque contemporaine que l'homme a commencé à comprendre les échelles immenses qui étaient en jeu. C'est avec le premier télescope géant de 2,50 m, installé au Mont Wilson, en Californie que l'idée des galaxies a vraiment vu le jour. L'américain Edwin Hubble a réussi à montrer alors que les nébuleuses spirales étaient des galaxies lointaines semblables à la nôtre et à établir la relation entre le décalage vers le rouge de leurs raies spectrales et leur distance. C'est avec lui que naît véritablement la cosmologie observationnelle. Après la Seconde Guerre mondiale de nombreux développements techniques sont réalisés qui permettent l'observation du ciel dans le domaine radio. En particulier, le rayonnement fossile découvert en 1965 dans le domaine des ondes millimétriques, se révèle être un outil extraordinairement puissant pour étudier notre Univers à grande échelle, notamment par les résultats des observations réalisées à l'aide des sondes COBE (1989) et WMAP (2003).

PLUS RÉCEMMENT, en 1980, les astronomes ont commencé à élaborer des catalogues de galaxies tridimensionnels profonds en commençant par le CfA1 (Huchra et al., 1988, 1992), le SDSS (Stoughton et al., 2002), le 2dF (Colless et al., 2003), le 6dFGS (Jones et al., 2004) et le 2MASS Redshift Survey (Huchra et al., 2005). Ces catalogues rivalisent en profondeur, bandes spectrales, nombre de galaxies et couverture angulaire sur le ciel. De nombreux résultats importants sur la cosmologie peuvent être extraits de ces catalogues. On peut citer par exemple des contraintes sur la statistique des fluctuations de matière à grande échelle, les paramètres cosmologiques, . . . . Néanmoins, ces résultats ne font qu'exploiter la distribution spatiale de galaxies. Or, le champ de vitesse de ces galaxies constitue une meilleure sonde pour l'étude des structures à grande échelle et l'identification des galaxies ou des groupes de galaxies générant le champ de gravité. Cette sonde présente d'autant plus d'intérêt que l'on sait d'après les expériences COBE et WMAP que le contenu énergétique de l'Univers est a priori composé d'environ 30% de matière, dont seulement 16% correspond à de la matière usuelle. On ne sait rien de la physique de cette matière noire si ce n'est qu'elle se comporte comme de la matière baryonique vis à vis du champ de gravité ; avoir une sonde capable d'indiquer quelle est sa distribution précise devrait apporter une aide précieuse. Malheureusement, le champ de vitesse est difficile (voire impossible) à mesurer directement. Nous allons donc devoir établir une stratégie d'étude du champ de vitesse passant par l'intermédiaire de la reconstruction des orbites des galaxies fondée seulement sur leur position actuelle.

CETTE THÈSE se propose de d'étudier ce qu'on appelle les reconstructions lagrangiennes du champ de vitesse à partir d'une distribution de masse. Nous en détaillerons les motivations théoriques, les tests pratiques sur des simulations et des catalogues virtuels, ainsi que l'application aux données observationnelles. Ce travail s'articule en cinq chapitres :

- Dans le premier chapitre, nous reviendrons sur les catalogues de galaxies et détaillerons quelques structures de notre Univers local afin de poser le contexte observationnel. Après ces quelques observations, nous reprendrons les concepts fondamentaux de la cosmologie gravitationnelle. Nous analyserons en premier lieu l'origine de l'intérêt porté aux champs de vitesse et quelques algorithmes existants qui permettent de les retrouver par inférence à partir de la distribution des galaxies à l'instant présent.

- Nous nous pencherons ensuite, dans le deuxième chapitre, sur une méthode de reconstruction particulière, appelée méthode de reconstruction de Monge-Ampère-Kantorovitch. Cette méthode sera testée sur une distribution de masse issue d’une simulation à  $N$ -corps et nous tenterons de retrouver le champ de vitesse et les conditions initiales de cette simulation à l’aide uniquement de la distribution finale de masse.
- Nous réfléchirons au chapitre trois sur deux méthodes de comparaison du champ de vitesse reconstruit au champ de vitesse observé, et ce qui peut être déduit concernant la densité moyenne de matière dans l’Univers.
- Nous analyserons au chapitre quatre à quels biais observationnels sont sensibles en général, les méthodes de reconstruction lagrangienne en les testant plus particulièrement à l’aide de la reconstruction MAK. Nous proposerons et testerons des moyens de corriger ces biais afin d’avoir une reconstruction fidèle du champ de vitesse tracé par les objets du catalogue (virtuels ou réels).
- Nous en viendrons enfin à l’analyse du “NearBy Galaxy catalog 3000 km s<sup>-1</sup>” (NBG-3k) et du 2MASS Redshift Survey (2MRS) et à la reconstruction du champ de vitesse de notre Univers local à partir de leurs données. Nous tenterons d’obtenir une mesure de l’origine de notre mouvement par rapport aux grandes structures, dont le référentiel de repos est identifié à celui du fond diffus cosmologique (Cosmic Microwave Background, ou CMB en anglais). Nous en viendrons ensuite à la mesure de la densité moyenne de notre Univers à l’aide de la comparaison entre vitesses reconstruites et observations.

## 2 Notre Univers : cosmographie et dynamique locale

### Sommaire

<b>2.1</b>	<b>Notre Univers Local / catalogues de galaxie.</b>	<b>3</b>
<b>2.2</b>	<b>L'Univers homogène</b>	<b>7</b>
2.2.1	Métrie de Friedmann-Lemaître-Robertson-Walker (FLRW)	7
2.2.2	Formes réduites – contenu de l'Univers	10
2.2.3	Quelques modèles d'Univers usuels	10
2.2.4	Distances dans l'Univers	11
2.2.5	Mesures les plus récentes des paramètres cosmologiques	13
2.2.6	Cinématique des objets dans un Univers de FLRW	13
<b>2.3</b>	<b>L'Univers inhomogène : dynamique linéaire des fluctuations</b>	<b>14</b>
2.3.1	L'approche Eulérienne	14
2.3.2	Théorie linéaire	15
2.3.3	L'approche Lagrangienne : l'approximation de Zel'dovich	17
<b>2.4</b>	<b>Intérêt et obtention des champs de vitesse</b>	<b>18</b>
2.4.1	Intérêt des champs de vitesse	18
2.4.2	Les distances dans l'Univers / Mesurer les vitesses	19
2.4.3	Les méthodes de reconstruction	21
2.4.4	La méthode POTENT	22
2.4.5	La méthode de moindre action	22
2.4.6	La reconstruction de Monge-Ampère-Kantorovitch (MAK)	26

DANS ce chapitre, nous allons commencer par un bref historique des catalogues de galaxies environnement (section 2.1). Nous exposerons ensuite la théorie classique de la dynamique de l'Univers homogène (section 2.2). Nous traiterons le cas de la dynamique des inhomogénéités dans un tel Univers (section 2.3) pour ensuite nous concentrer sur leur lien avec les champs de vitesse (section 2.4). Nous verrons quelles méthodes sont utilisées pour tenter d'observer le champ de vitesse de l'Univers local (section 2.4.2). Nous nous pencherons enfin sur quelques méthodes de reconstruction (POTENT, Fast-Action-Method, et Monge-Ampère-Kantorovitch).

### 2.1 Notre Univers Local / catalogues de galaxie.

Depuis les années 1980, les astronomes font des campagnes de mesure des vitesses radiales et des magnitudes des galaxies dans différentes régions du ciel. Ces données sont réduites et rassemblées dans des bases de données appelées catalogues de galaxies. Le premier a été le catalogue LICK, en 1967 (Shane & Wirtanen, 1967). Il consistait en un relevé des positions angulaires des galaxies dans le ciel, les données tridimensionnelles tant inaccessibles. Une image de la partie dans l'hémisphère galactique nord, c'est-à-dire la portion de l'Univers se trouvant du côté nord du plan de la Voie Lactée, est donnée figure 2.1. Les deux faces de ce plan sont orientées de manière à ce que :

- le pôle nord galactique corresponde à la direction de la constellation de la Chevelure de Bérénice, près de l'étoile  $\alpha$  Bootis (Arcturus) ;
- le pôle sud galactique corresponde à la direction de la constellation du Sculpteur.

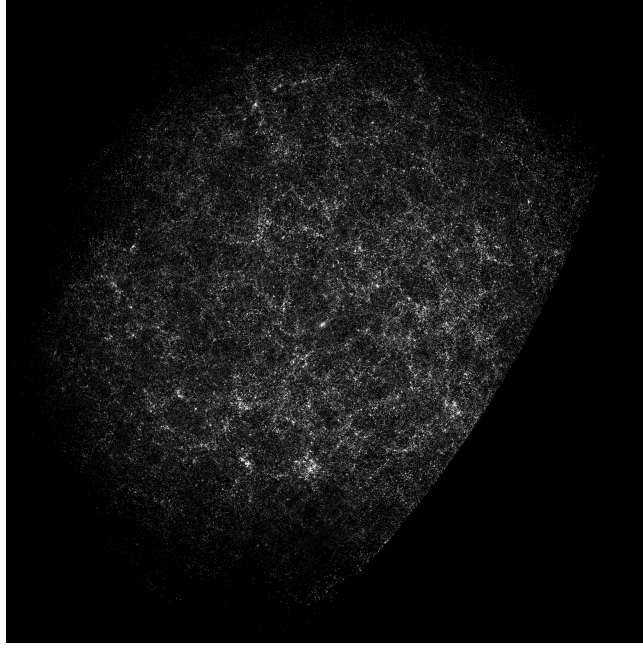


FIG. 2.1: *Catalogues de galaxies* – image du catalogue LICK, un des tout premier catalogues de galaxies ne recensant que les coordonnées angulaires de celle-ci. Cette image provient du site <http://groth2005.princeton.edu/~groth>.

Le premier catalogue “tridimensionnel”, avec mesure de décalage spectraux, a été le CfA1 (Huchra et al., 1988, cf. également figure 2.2). On a alors pour la première fois mis en évidence la structuration à grande échelle des galaxies, et on a notamment découvert le “Grand Mur”. Au cours des vingt dernières années, on a vu une explosion du nombre et de la diversité des catalogues existants. Néanmoins chaque catalogue a des propriétés bien spécifiques de couverture angulaire, de nombres de galaxies détectées, de profondeur en distance, de limite en magnitude et/ou longueur d’onde. On peut décrire succinctement quelques avantages et inconvénients de quelques catalogues majeurs :

- **Sloan Digital Sky Survey (SDSS 1,2,3)** <http://www.sdss.org/> (Pour les premières données publiques Stoughton et al., 2002).

C’est un catalogue profond en redshift ( $z \lesssim 0.3$  pour l’échantillon principal et  $0.3 \lesssim z \lesssim 0.5$  pour l’échantillon des galaxies lumineuses dans le rouge), limité dans le rouge ( $r < 17.7$ ), avec une bonne couverture angulaire ( $\sim 15\%$  du ciel). On compte alors 1 271 680 spectres observés (DR6). Parmi ceux-ci on a mesuré le redshift d’un demi million de galaxies (avec une précision de  $30 \text{ km s}^{-1}$ ). Néanmoins, la stratégie d’observation est compliquée et le catalogue de redshift n’est encore réalisé que par morceaux sur le ciel. Selon les méthodes que l’on souhaite appliquer sur les données, il peut donc y avoir de considérables effets de bord. Par ailleurs, l’échantillon principal a été sélectionné en bande photométrique R, qui ne trace pas aussi bien que la bande K la masse stellaire.

- **Six Degree Field Galaxy Survey (6dFGS)** <http://www.aao.gov.au/local/www/6df/> (Jones et al., 2004, 2005).

C’est un catalogue moins profond que le SDSS mais qui a l’avantage de couvrir un demi-ciel. Les galaxies ont été sélectionnées avec le 2MASS, donc en bande K (proche infrarouge). L’échantillon principal de galaxie est issu du catalogue 2MASS. On a sélectionné les galaxies dont la magnitude  $K_S \leq 12.75$ . Au total, 89 211 redshifts sont disponibles dans la base de données publique (au mois d’Avril 2008). La médiane de la distribution en redshifts des galaxies est  $\bar{z} \simeq 0.054$ . Le relevé évite la région du ciel de latitude galactique  $|b| < 10$  degrés à cause de l’absorption de la lumière par le plan galactique de la Voie Lactée. Ce catalogue de redshifts devrait être couplé dans un

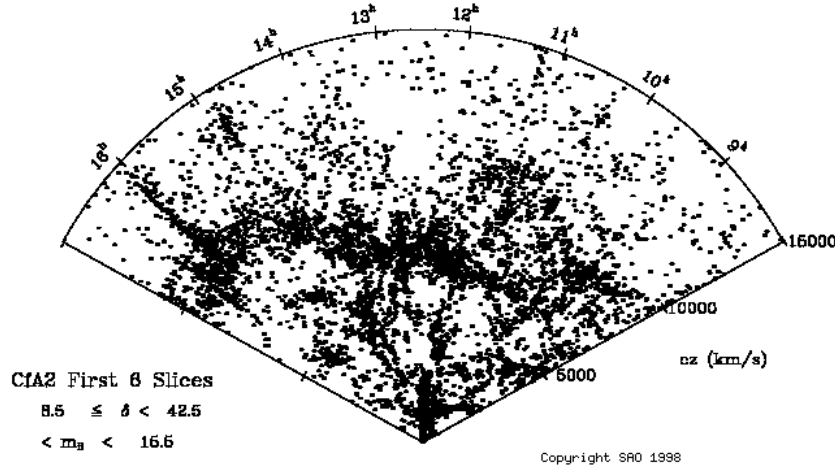


FIG. 2.2: *Catalogues de galaxies* – Une coupe du catalogue CfA2, successeur du CfA1, où l’on voit clairement la présence d’un “mur” de galaxies entre 8h et 17h et entre 5000 et 10000 km s<sup>-1</sup>. Chaque galaxie est représentée par un point. L’image est tirée du site <http://cfa-www.harvard.edu/~huchra/zcat/>.

futur proche à un catalogue de distances mesurées par la méthode du plan fondamental (§ 2.4.2) sur le sous-ensemble des galaxies dominés par le bulbe.

- **2Mass Redshift Survey** <http://cfa-www.harvard.edu/~huchra/2mass/> (Huchra et al., 2005) C’est un catalogue dont les galaxies ont été elles-aussi sélectionnées avec le 2MASS, avec  $K \leq 11.25$ . Son énorme avantage est d’avoir une couverture complète du ciel mis à part le plan galactique. Il a une profondeur médiane de  $z = 0.02$  plutôt faible par rapport au SDSS et au 6dFGS et un nombre plus réduit de galaxies. Néanmoins il ne s’agit pour le moment que de la première version du catalogue et le but ultime est d’avoir une limite de magnitude de  $K \leq 12.75$ , ce qui en fera le catalogue le plus profond et avec une couverture du ciel quasi parfaite. Les données de ce catalogue ne sont pas publiques. Nous en reparlerons de manière plus détaillée au chapitre 6.

Dans les années à venir, nous devrions nous attendre à une explosion de nouvelles données disponibles et de nouveaux catalogues par l’intermédiaire des grands relevés qui seront effectués par, par exemple, PANSTARRS (<http://pan-starrs.ifa.hawaii.edu/>), le SDSS-3 (<http://www.sdss.org>), ou plus lointain le projet SPACE. Ce dernier projet, non encore validé par l’ESA, est de loin le plus ambitieux car il vise à mesurer la position de 500 millions de galaxies jusqu’à  $z \sim 2$  (position angulaire et décalage spectral).

La distribution des galaxies dans l’Univers local est néanmoins maintenant assez bien connue en dehors du plan galactique de la Voie Lactée. On constate que les galaxies s’assemblent en des structures plus grosses qu’on appelle amas (un groupe de galaxie assez massif), superamas (un ensemble de groupes/filament de galaxies) et mur (un plan épais de galaxies). Quelques unes des structures les plus importantes sont résumées dans la table 2.1 et représentées schématiquement figure 2.3.

Nom	Type	l	b	$\sim z/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$
Puppis	Amas	240	-28	10
Virgo	Amas	282	74	10
Hydra	Amas	270	50	30
Centaurus	Amas	300	20	35
Pavo-Indus-Telescopium	Amas	0	-45	55
C3	Amas	275	-30	50
C5	Amas	195	0	50
C6	Amas	170	-15	50
C7	Amas	55	15	50
Cancer	Amas	210	30	50
Cetus	Amas	190	-55	50
Camelopardalis	Amas	145	30	50
Norma	Amas	330	-10	50
Coma	Amas	58	87	70
Shapley	Super-amas	306	30	130
Hercules	Super-amas	30	43	100
Perseus-Pegasus	Super-amas	95	-42	120
Ophiucus	Super-amas	0	10	85
Lepus	Super-amas	233	-24	120
Horologium-Reticulum	Super-amas	261	-54	160
Pisces-Cetus	Super-amas	95	-82	160
Pegasus	Mur	90	-5	50
Perseus-Pisces	Mur	160	-5→-30	50
C2	Surdensité	345	10	35
C8	Surdensité	295	7	35
Centaurus-Crux	Amas	295	7	
C4	Surdensité	240	25	35
Hydra	Vide	285	30	40
Sculptor	Vide	300	-15	60
Local Void	Vide	0	0	0→50
Great Attractor		310	20	20→60

TAB. 2.1: *Catalogues de galaxies* – Une liste de quelques unes de structures les plus importantes de notre Univers local.

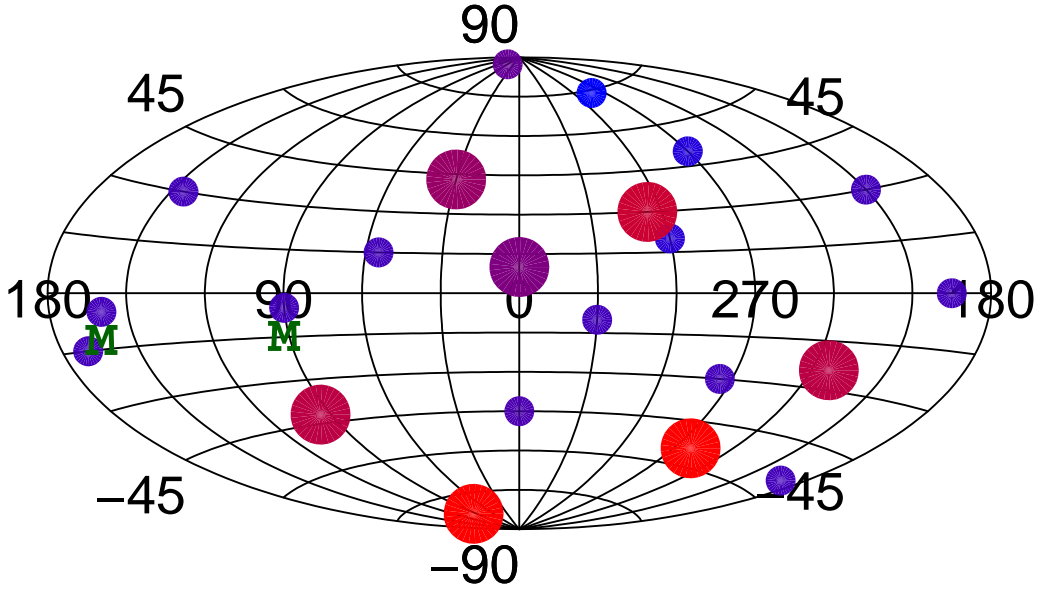


FIG. 2.3: *Cosmographie*– Ce graphique représente quelques structures importantes de notre Univers local. Ces structures ont été listées dans la table 2.1. Les disques représentent les amas de galaxies. Les gros disques représentent les super-amas de galaxies. Un *M* dénote la présence d'un mur de galaxies. Le superamas de Shapley est particulièrement étudié car il représente une des plus importantes concentration de galaxies de l'Univers local. D'après par exemple Ettori et al. (1997), il contient une masse pouvant être de l'ordre  $10^{16} M_{\odot}$ .

Nous allons à présent nous attacher à trouver un modèle capable d'expliquer comment sont apparues de telles structures, comment les galaxies se forment, quels sont les paramètres fondamentaux de ce système en évolution appelé Univers. Nous allons expliciter à présent les modèles “standards” de la cosmologie, en premier lieu le modèle de l'Univers homogène (§ 2.2), puis comment des petites fluctuations de densité de matière peuvent évoluer et engendrer de telles structures (§ 2.3). Ensuite nous nous pencherons sur l'objet principal de cette thèse qui est la reconstruction des vitesses propres des galaxies (§ 2.4).

## 2.2 L'Univers homogène

Nous allons ici présenter les bases de la cosmologie moderne qui se fondent sur des modèles dynamique de l'Univers dont l'évolution est régie par un nombre restreint de paramètres. Le but ici n'est pas de reprendre en détail ce qui existe dans les ouvrages de références tels que Thorne et al. (1973); Peebles (1980); Wald (1984); Uzan & Peter (2005) mais d'introduire les concepts fondamentaux d'une manière concise pour motiver les méthodes que nous allons aborder dans cette thèse.

### 2.2.1 Métrique de Friedmann-Lemaître-Robertson-Walker (FLRW)

L'approche moderne de la cosmologie repose principalement sur le *principe cosmologique*, ou sa version plus faible le *principe copernicien*. Le premier stipule que l'Univers dans son ensemble est isotrope et homogène. Il s'agit d'une hypothèse très forte qui revient à déterminer la géométrie totale de l'Univers à partir d'une observation locale. On peut lui opposer le second principe pour qui l'observateur n'occupe pas un emplacement privilégié dans l'univers, celui-ci tant par ailleurs isotrope. Chacun de ces deux principes nous amène en pratique au même résultat. Cependant, si on utilise le principe copernicien et non le principe cosmologique, la solution obtenue ne sera vraie que pour l'Univers observable et pas au-delà. Nous allons supposer qu'il existe une échelle spatiale à partir de



laquelle le principe cosmologique est vrai. En dessous de cette échelle, la dynamique des fluctuations sera traitée perturbativement (cf. section 2.3) par rapport à la dynamique d'ensemble de l'Univers.

Pour décrire cette dynamique, nous allons nous appuyer sur la relativité générale qui est, jusqu'à présent, la seule bonne théorie capable de décrire les phénomènes gravitationnels à grande échelle dans le domaine des champs forts. La relativité générale est basée sur l'étude de la dynamique d'un objet appelé métrique. Cet objet permet de mesurer les distances dans l'Univers. Pour un espace euclidien, on peut écrire la distance élémentaire entre deux points  $\mathbf{x}(x, y, z)$  et  $\mathbf{y}(x + dx_1, y + dx_2, z + dx_3)$  comme :

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 = \sum_{i,j=1}^3 \eta_{i,j} dx_i dx_j = \eta_{i,j} dx_i dx_j \quad (2.1)$$

avec  $\eta_{i,j} = \delta_{i,j}$ ,  $\delta_{i,j}$  étant le symbole de Kronecker. Dans l'équation ci-dessus, on introduit aussi la notation dite d'*Einstein* où on omet le signe somme sur un indice quand celui-ci est répété. Déjà, la théorie de la relativité restreinte nous enseigne qu'une bonne métrique ne pourra de toute façon pas séparer le temps de l'espace. Ceux-ci doivent être mis sur un pied d'égalité et que les distances doivent être mesurées dans un sur-espace, appelé de *Minkowski*, où cette fois les distances s'écrivent

$$ds^2 = c^2 dt^2 - \eta_{i,j} dx_i dx_j = \eta_{\mu,\nu} dx_\mu dx_\nu \quad (2.2)$$

avec

$$\eta_{\mu,\nu} = \begin{cases} c^2 & \text{si } \mu = \nu = 0 \\ -1 & \text{si } \mu = \nu \neq 0 \\ 0 & \text{sinon} \end{cases}, \quad (2.3)$$

et  $c$  la vitesse de la lumière dont la valeur est fixée par le bureau international des poids et mesures à  $c = 299\,792\,458 \text{ m s}^{-1}$ . Le tenseur  $\eta_{\mu,\nu}$  est la métrique de l'espace-temps de Lorentz.

La relativité générale va encore plus loin en considérant que l'objet métrique, écrit  $g_{\mu,\nu}$ , est lui-même totalement dynamique. A partir d'arguments généraux basés sur le principe d'Einstein, ou principe d'équivalence faible, ou bien encore l'Universalité de la chute libre, il est possible de montrer que la métrique doit obéir à l'équation dite d'Einstein

$$R_{\mu,\nu} - \frac{1}{2} R g_{\mu,\nu} = \frac{8\pi G}{c^4} T_{\mu,\nu} + \Lambda g_{\mu,\nu}, \quad (2.4)$$

avec  $R_{\mu,\nu}$  le tenseur de Ricci de la métrique  $g_{\mu,\nu}$ ,  $R = R_{\mu,\nu} g^{\mu,\nu}$  le scalaire de courbure,  $T_{\mu,\nu}$  le tenseur énergie-impulsion décrivant le contenu de l'Univers. Le membre de droite est souvent regroupé dans un tenseur, dit d'Einstein, noté  $G_{\mu,\nu}$ . Le tenseur  $G_{\mu,\nu}$  ne dépend au final que de la métrique et de ses dérivées. Cette équation relie donc le géométrique (la métrique) au contenu matériel (le tenseur énergie-impulsion). A cette équation nous avons ici introduit la constante cosmologique  $\Lambda$  dont nous discuterons l'impact dans une prochaine sous-section. Son statut intrinsèque est ambigu car elle peut être considérée soit comme un effet géométrique soit comme une énergie du vide suivant que l'on place le terme respectivement à gauche ou à droite dans l'équation ci-dessus.

A partir des symétries induites par le principe cosmologique, on peut en fait simplifier grandement le problème. Robertson et Walker ont montré que, sous cette hypothèse, on peut déduire que la métrique doit obligatoirement prendre la forme suivante (dite métrique de Friedmann-Lemaître-Robertson-Walker)

$$ds^2 = g_{\mu,\nu} dx_\mu dx_\nu = -c^2 dt^2 + a^2(t) \gamma_{i,j} dx_i dx_j = -c^2 dt^2 + a^2(t) \left( \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right) \quad (2.5)$$

avec  $t$  le temps cosmique,  $a(t)$  le facteur d'échelle,  $k$  le facteur de courbure et  $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$  la métrique de la sphère de dimension deux. Le facteur d'échelle  $a(t)$  décrit l'expansion de l'Univers à un temps  $t$ . Le facteur de courbure peut prendre une valeur positive (pour un Univers à courbure positive, dit "fermé"), nulle (Univers plat) ou négative (Univers à courbure négative, dit "en selle de

cheval"). On peut alors montrer que le tenseur de Ricci de la métrique de FLRW est diagonal et que ses éléments diagonaux valent

$$R_{0,0} = -3\frac{\ddot{a}}{a} \quad (2.6)$$

$$R_{i,j} = \left(2H + \frac{\ddot{a}}{a} + 2k\right) a^2 \gamma_{i,j} . \quad (2.7)$$

avec  $H = \frac{\dot{a}}{a}$ . On peut en déduire le scalaire de courbure

$$R = 6 \left( H^2 + \frac{\ddot{a}}{a} + k \right) \quad (2.8)$$

et enfin le tenseur d'Einstein, qui est lui aussi diagonal de par l'équation (2.4)

$$G_{0,0} = 3(H^2 + k) \quad G_{i,j} = -(H^2 + \frac{\ddot{a}}{a} + k) a^2 \gamma_{i,j} . \quad (2.9)$$

### Equations de Friedmann

Afin de résoudre l'équation d'Einstein, il ne nous reste plus qu'à définir le tenseur énergie-impulsion. A partir des symétries de l'espace-temps, on peut montrer que la forme la plus générale de ce tenseur est celle d'un fluide parfait :

$$T_{\mu,\nu} = \rho u_\mu u_\nu + P(u_\mu u_\nu - g_{\mu,\nu}) \quad (2.10)$$

avec  $u_\mu = dX_\mu/ds$  la quadri-vitesse du contenu de l'Univers. Dans le repère où ce fluide est au repos, on aura donc  $T_{\mu,\nu} = \text{diag}(\rho, P, P, P)$ . La seule liberté restante dans ce tenseur est l'expression de l'équation d'état reliant  $P$  et  $\rho$ . En utilisant les tenseurs (2.9) et (2.10) on trouve deux équations indépendantes appelées équations de Friedmann :

$$H^2 = \frac{8\pi G}{3c^4} \rho - \frac{K}{a^2} + \frac{\Lambda}{3} \quad (2.11)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} (\rho + 3P) + \frac{\Lambda}{3} . \quad (2.12)$$

La conservation du tenseur énergie-impulsion ( $\partial_\mu T^{\mu,\nu}$ ) se réduit, elle, à une seule équation :

$$\dot{\rho} + 3H(\rho + P) = 0 . \quad (2.13)$$

Les deux équations de Friedmann et l'équation de conservation ci-dessus ne sont en fait pas indépendantes. On peut s'en convaincre après quelques manipulations algébriques sur ces trois équations. L'équation de conservation peut aussi se ré-écrire comme :

$$\frac{d(\rho a^3)}{dt} = -P \frac{da^3}{dt} , \quad (2.14)$$

avec  $a(t)$  le facteur d'expansion défini précédemment. On a donc deux équations indépendantes et trois inconnues. Il nous faut postuler une relation supplémentaire. Généralement on fixe la forme de l'équation d'état pour qu'elle prenne une forme simple :

$$P = w\rho . \quad (2.15)$$

Suivant le type des "particules" composant le fluide considéré on aura un  $w$  différent. Par exemple :

- pour de la matière ordinaire, on a  $\rho \propto a^{-3}$ , donc d'après l'équation (2.14) on a  $P = 0$  et donc  $w = 0$ .
- $P = \rho/3$  et donc  $w = 1/3$  pour des particules relativistes comme des photons.
- $w = -1/3$  pour le terme de courbure  $K$ .
- $w = -1$  pour la constante cosmologique  $\Lambda$ .

### 2.2.2 Formes réduites – contenu de l'Univers

On écrit en général les différentes densité  $\rho_X$  sous une forme réduite, adimensionnée, en divisant par la densité critique

$$\rho_c = \frac{3H^2}{8\pi G} . \quad (2.16)$$

On définit alors les quantités réduites :

$$\Omega = \frac{\rho}{\rho_c} \quad \Omega_K = -\frac{K}{H^2 a^2} \quad \Omega_\Lambda = \frac{\Lambda}{3H^2} \quad (2.17)$$

Traditionnellement, on sépare  $\Omega$  en une somme de composantes ayant des équations d'état différentes :

$$\Omega = \sum_X \Omega_X \quad (2.18)$$

et nous noterons  $\Omega_{X,0}$  sa valeur à l'instant présent. On a résumé les composantes qui ont été identifiées à ce jour dans la liste suivante :

$$\text{Matière-énergie } (\Omega_0) \left\{ \begin{array}{ll} \text{Matière ordinaire } (\Omega_m) & \left\{ \begin{array}{l} \text{Matière baryonique } (\Omega_b) \\ \text{Matière noire } (\Omega_m) \end{array} \right. \\ \text{Particules relativistes } (\Omega_r) & \left\{ \begin{array}{l} \text{Photons } (\Omega_\gamma) \\ \text{Neutrinos } (\Omega_\nu) \end{array} \right. \\ \text{Constante cosmologique } (\Omega_\Lambda) & \end{array} \right.$$

On peut alors réécrire la première équation de Friedmann comme une contrainte :

$$\boxed{\left( \sum_X \Omega_X \right) + \Omega_\Lambda + \Omega_K = 1 .} \quad (2.19)$$

On définit pour chacun un  $w_X$  tel que  $P_X = w_X \rho_X$ , ce qui nous permet d'écrire l'équation d'évolution de sa densité en fonction du paramètre d'échelle et de la constante de Hubble

$$\Omega_X = \Omega_{X,0} \left( \frac{a}{a_0} \right)^{-3(1+w_X)} \left( \frac{H_0}{H} \right)^2 , \quad (2.20)$$

avec  $H_0$  la constante de Hubble  $H$  évaluée à l'instant présent.

### 2.2.3 Quelques modèles d'Univers usuels

Imaginons un instant que l'Univers ne contiennent que de la matière ordinaire ( $\Omega_r = \Omega_\Lambda = 0$  ci-dessus) . Il est alors possible de décrire son évolution en fonction de la valeur de  $\Omega_0 = \Omega_{m,0}$  seul. En effet à tout instant  $\rho_c = \rho$  d'après la première équation de Friedmann (2.11). Nous allons ici définir de plus une quantité, appelé redshift et notée  $z$ , dont on fera le lien avec le décalage spectral vers le rouge dans une prochaine section :

$$1 + z = \frac{a_0}{a} , \quad (2.21)$$

avec  $a_0$  l'échelle de l'Univers à l'instant présent. On peut résoudre analytiquement l'évolution de l'Univers en fonction de  $\Omega_{m,0}$  et trouver la relation liant son échelle (ou de manière équivalente  $z$ ) à son âge :

– si  $\Omega_{m,0} = 1$ ,

$$t = \frac{2}{3H_0(1+z)^{3/2}} . \quad (2.22)$$

– si  $\Omega_{m,0} > 1$ ,

$$t = \frac{\Omega_{m,0}}{2H_0 (\Omega_{m,0} - 1)^{3/2}} \left( \cos^{-1} \left[ \frac{2 + \Omega_{m,0}(z - 1)}{(1 + z)\Omega_{m,0}} \right] - \frac{2\sqrt{(1 + z\Omega_{m,0})(\Omega_{m,0} - 1)}}{(1 + z)\Omega_{m,0}} \right) \quad (2.23)$$

– si  $0 \leq \Omega_{m,0} < 1$ ,

$$t = \frac{\Omega_{m,0}}{2H_0 (1 - \Omega_{m,0})^{3/2}} \left( -\cosh^{-1} \left[ \frac{2 + \Omega_{m,0}(z - 1)}{(1 + z)\Omega_{m,0}} \right] + \frac{2\sqrt{(1 + z\Omega_{m,0})(1 - \Omega_{m,0})}}{(1 + z)\Omega_{m,0}} \right) \quad (2.24)$$

Einstein a alors tout de suite remarqué qu'on ne parviendrait pas à obtenir une solution statique en ne supposant un Univers rempli que de matière. Dans le premier cas, l'Univers est plat ( $\Omega_K = 0$ ) et est en expansion éternelle ( $a \propto t^{2/3}$ ). Dans le deuxième cas, l'Univers est sphérique ( $\Omega_K < 0$ ) et finit par s'effondrer sur lui-même, le facteur d'expansion  $a$  décrivant une sorte de cycloïde en fonction du temps. Dans le troisième cas, l'Univers a une géométrie hyperbolique ( $\Omega_K > 0$ ) a aussi une expansion éternelle. Comme le paradigme d'alors était de considérer que l'Univers était statique, Einstein découvrit qu'il pouvait introduire une constante  $\Lambda$  telle que la dynamique de l'Univers soit figée ( $H = 0$ ). Mais à la même époque, Edwin Hubble découvrit que les galaxies qu'on observait avaient toutes un décalage vers le rouge en moyenne proportionnel à leur distance

$$v_{\text{galaxie}} \simeq H_0 d, \quad (2.25)$$

avec  $v_{\text{galaxie}}$  la vitesse apparente de la galaxie obtenue par l'observation de son spectre,  $d$  la distance estimée de cette galaxie,<sup>1</sup> et  $H_0 \simeq 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$  actuellement. Cette mesure remet en cause l'hypothèse d'un Univers statique que les cosmologistes retenaient à cette époque. Cette constante fut alors abandonnée jusqu'au début des années 90 où l'on fit les premières mesures de distance d'objet situés à distance cosmologique : les supernovae. On découvrit que si l'espace est plat ( $\Omega_K = 0$ ) alors on doit avoir des paramètres cosmologiques de l'ordre

$$\begin{aligned} \Omega_{m,0} &= 0.28 \pm 0.14 \\ \Omega_{\Lambda,0} &= 1 - \Omega_{m,0} = 0.72 \pm 0.14 \end{aligned} \quad (2.26)$$

En fait, cette hypothèse d'Univers plat est très bien supportée par les mesures de WMAP5 qui donnent une borne à la courbure  $\Omega_K = -0.099^{+0.085}_{-0.100}$ .

## 2.2.4 Distances dans l'Univers

### Décalage vers le rouge

Nous n'avions pas montré la relation (2.21) dans la section précédente. Nous allons, pour réaliser cela, utiliser la métrique de FLRW (2.5). On suppose qu'un photon est émis à l'instant  $t_1$  par un objet immobile à une distance  $r_1$  de l'observateur placé à l'origine. Ce photon va se déplacer purement radialement pour des raisons de symétrie et être détecté à un instant  $t_0$  par l'observateur. Comme  $ds^2 = 0$  pour un photon, on va avoir, sur la géodésique empruntée par le photon,

$$c^2 \int_{t_1}^{t_0} \frac{dt}{a(t)} = - \int_{r_1}^0 \frac{dr}{\sqrt{1 - kr^2}}, \quad (2.27)$$

d'après (2.5). Comme cet objet est immobile on peut en particulier écrire que, pour deux photons émis respectivement à  $t_1$  et  $t_1 + \delta t_1$ , avec  $\delta t_1$  infiniment petit :

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{t_1 + \delta t_1}^{t_0 + \delta t_0} \frac{dt}{a(t)} = \int_{t_1}^{t_0} \frac{dt}{a(t)} + \frac{\delta t_0}{a(t_0)} - \frac{\delta t_1}{a(t_1)}, \quad (2.28)$$

<sup>1</sup>L'estimation de la distance intrinsèque d'un objet sera traité dans une section suivante.

ce qui nous amène à l'égalité

$$\frac{\delta t_0}{a(t_0)} = \frac{\delta t_1}{a(t_1)} \quad (2.29)$$

reliant le rythme auquel sont reçus les photons par l'observateur à celui auquel ils sont envoyés par la source. Le rapport de ces deux temps est l'inverse du rapport des longueurs d'onde des photons émis et reçus. On a donc

$$\frac{\lambda_1}{\lambda_0} = 1 + z = \frac{a(t_0)}{a(t_1)} = \frac{a_0}{a} \quad (2.30)$$

Il nous faut à présent relier cette distance à la “vraie” distance de l'objet.

### Distance angulaire

Un objet de taille propre  $D$  donnée sera vu sous un angle  $\alpha$ , si cet angle est petit, alors on peut définir la distance  $d_A$  par

$$\alpha = \frac{D}{d_A} . \quad (2.31)$$

Si on écrit  $ds^2 = 0$  pour la distance transversale d'un objet, donc à  $x_1$  fixe, on a  $dt = a(t_1)x_1(t_1)\alpha$  sur la géodésique du photon en utilisant la métrique (2.5). On a donc simplement

$$d_A = a(t_1)x_1(t_1) . \quad (2.32)$$

On peut réécrire la relation (2.27) avec le changement de variable (2.30)

$$\int_0^{z_1} \frac{cdz}{a_0 H(z)} = \int_0^{x_1} \frac{dx}{\sqrt{1 - kx^2}} \quad (2.33)$$

A l'ordre 0, pour des objets peu éloignés, on a  $H(z) \simeq H(z = 0) = H_0$  donc la relation ci-dessus devient

$$cz_1 = a_0 H_0 x_1 = \frac{a_0}{a_1} H_0 d_A \simeq H_0 d_A . \quad (2.34)$$

On a ainsi relié le redshift  $z_1$  à la distance angulaire  $d_A$  d'un objet peu éloigné de nous. Pour des objets à plus grand redshift il faudrait tenir compte de l'histoire complète de la métrique, ce qui ferait intervenir les paramètres cosmologiques  $\Omega_X$ . On peut remarquer que cette relation est très proche de la loi empirique de Hubble décrite par l'équation (2.25).

### Distance lumineuse

Une autre distance peut être construite à partir des observations. Si on suppose qu'on est capable de connaître la luminosité absolue  $L$  d'un objet, on peut comparer celle-ci au flux  $l$  observé sur Terre et en déduire une distance  $d_L$  :

$$d_L = \sqrt{\frac{L}{4\pi l}} . \quad (2.35)$$

Dans un espace Euclidien, une onde sphérique émise se dilue avec le carré de la distance parcourue par les photons et  $d_L$  correspond alors à la distance usuelle. Dans l'Univers en expansion, les choses sont cependant différentes. L'énergie émise pendant un temps  $\delta t_1$  est reçue en un temps  $\delta t_0 = (1 + z)\delta t_1$  mais cette énergie est de plus divisée par un facteur  $1 + z$  du fait de l'augmentation de la longueur d'onde des photons. Pour une source se trouvant à une distance  $d_1$  de l'observateur au moment de l'émission, on obtient donc :

$$l = \frac{L\delta t_1}{\delta t_0(1 + z)} \frac{1}{4\pi d_1^2 (a_0/a_1)^2} = \frac{L}{4\pi x_1^2 a_0^2 (1 + z)^2} , \quad (2.36)$$

soit en utilisant (2.35) :

$$d_L = x_1 \frac{a_0}{a_1} (1 + z) = \frac{d_A}{a_1} (1 + z)^2 . \quad (2.37)$$

Parameter	Valeur moyenne
$\sigma_8$	$0.796 \pm 0.036$
$\Omega_m$	$0.258 \pm 0.030$
$\Omega_m h^2$	$0.1326 \pm 0.00063$
$H_0$	$71.9^{+2.6}_{-2.7} \text{ km s}^{-1} \text{ Mpc}^{-1}$
$z_{\text{reion}}$	$11.0 \pm 1.4$

TAB. 2.2: Les paramètres cosmologiques à l'heure actuelle tels que mesurés par WMAP5 (Dunkley et al., 2008).

### 2.2.5 Mesures les plus récentes des paramètres cosmologiques

D'après les analyses des observations du fond diffus cosmologique par la sonde WMAP (Spergel et al., 2006) et du spectre de puissance des fluctuations de la densité de galaxies dans le SDSS (Tegmark et al., 2006), nous savons à présent que la matière ne représente pas beaucoup plus qu'un tiers de l'énergie de masse totale de l'Univers. Les valeurs précises données par ces expériences pour les paramètres cosmologiques sont données table 2.2.

### 2.2.6 Cinématique des objets dans un Univers de FLRW

Ici nous allons nous intéresser à la cinématique d'objets plongés dans un espace dont la métrique est de type FLRW. Il est souvent agréable de décrire la distribution de matière ainsi que ses mouvements relativement au modèle homogène et isotrope. Avec un tel modèle, la séparation physique de deux particules varie avec le temps comme :

$$\mathbf{r} = a(t)\mathbf{x} \quad (2.38)$$

où  $\mathbf{x}$  est constant pour la paire et  $a(t)$  est le facteur d'expansion donné par la métrique (2.5).  $\mathbf{x}$  est parfois appelé le vecteur position comobile. Si on dérive par rapport au temps cette expression on trouve

$$\dot{\mathbf{r}} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}} \quad (2.39)$$

$$= H\mathbf{r} + a\dot{\mathbf{x}} \quad (2.40)$$

avec  $H$  la constante de Hubble. Si les particules sont immobiles on devrait bien sur avoir  $\dot{\mathbf{x}} = 0$ , mais ici on va s'intéresser aux déviations faibles par rapport à l'expansion de l'Univers qui est donné par le premier terme  $H\mathbf{r}$ . On définit ainsi ce qu'on appelle la vitesse propre  $\mathbf{u}$  d'une particule ou d'un objet comme :

$$\mathbf{u} \equiv a\dot{\mathbf{x}} = \dot{\mathbf{r}} - H\mathbf{r} . \quad (2.41)$$

Maintenant il nous faudrait avoir une description dynamique de l'évolution de ce vecteur vitesse à partir du contenu légèrement inhomogène de l'Univers. C'est l'objet de la section suivante. Si on revient sur la définition de la distance redshift, on voit que nous allons être vite confronté à un problème si on tente d'utiliser le redshift comme indicateur absolu de distance. A petit redshift, nous allons être aussi sensible aux mouvements des masses relativement par rapport à l'observateur. Cela veut dire également que la relation de Hubble doit être modifiée et le décalage vers le rouge  $z$  est en fait dans l'Univers local :

$$cz \simeq Hd + \mathbf{u} \cdot \hat{\mathbf{x}} \quad (2.42)$$

avec  $\hat{\mathbf{x}}$  le vecteur unitaire dans la direction de l'objet considéré,  $d$  sa distance (lumineuse ou angulaire) et  $\mathbf{u}$  sa vitesse propre.

## 2.3 L'Univers inhomogène : dynamique linéaire des fluctuations

Dans toute la suite de ce travail, nous allons appeler coordonnées lagrangiennes (généralement notées  $\mathbf{q}$  ici) d'une particule de matière, les coordonnées qu'auraient eu les particules de matière à  $z \rightarrow +\infty$  si les fluctuations du champ de gravité à tout instant sont données uniquement par la présence de matière. C'est une notation pratique car, à toute particule de matière noire, on peut affecter une et seulement une coordonnée lagrangienne. Au contraire, on appelle coordonnées eulériennes d'une particule de matière les coordonnées correspondant à sa position à un instant donné  $t$ . A une coordonnée eulérienne correspond en général un nombre indéfini de particules de matière. Nous allons à présent présenter deux formalismes approchés de la dynamique s'appuyant soit sur les coordonnées eulériennes (section 2.3.1) soit sur les coordonnées lagrangiennes (section 2.3.3).

### 2.3.1 L'approche Eulérienne

La théorie de croissance linéaire des structures est connue depuis les années 70 et de nombreux ouvrages font référence en la matière, citons principalement Peebles (1980) dans lequel tous les calculs sont très détaillés. Le lecteur intéressé est invité à se référer à cet ouvrage.

Nous allons supposer que la métrique de fond décrivant l'évolution globale de l'Univers est donnée. On suppose de plus que cette métrique est de type FLRW et que le paramètre d'échelle  $a(t)$  est donné. Nous allons partir de la distribution de particules de masse identique dans l'espace des phases :

$$dN = f(\mathbf{x}, \mathbf{p}, t) d^3\mathbf{x} d^3\mathbf{v}. \quad (2.43)$$

Une première équation à laquelle est soumise la fonction de distribution  $f$  est l'équation de Boltzmann non-collisionnelle, qui est une version locale du théorème de Liouville où le volume de l'espace des phases se conserve dans le temps :<sup>2</sup>

$$\frac{df}{dt} = 0 \Leftrightarrow \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{x}} \cdot \frac{\mathbf{p}}{ma^2} - m \mathbf{grad}_{\mathbf{x}} \Phi \cdot \frac{\partial f}{\partial \mathbf{p}} = 0, \quad (2.44)$$

avec  $\Phi$  le potentiel gravitationnel solution de l'équation de Poisson correspondant au contenu relatif en masse de l'Univers :

$$\Delta \Phi = 4\pi G a^2 \bar{\rho}(t) \delta(\mathbf{x}, t), \quad (2.45)$$

avec

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}. \quad (2.46)$$

Nous allons esquisser dans la suite le modèle basé sur cette équation qui est celui du *fluide* sans pression de matière noire. Ce modèle est bien sur une approximation de la description complète de l'état de l'ensemble des particules de matière noire qui nécessiterait de résoudre exactement l'équation (2.44). Cependant, elle donne déjà de nombreux résultats intéressants et utiles pour résoudre la plupart des problèmes à grande échelle. L'approche fluide permet de se limiter à étudier les deux premiers moments de l'équation (2.44) :

$$\rho(\mathbf{x}, t) = \int_{\mathbf{p}} d^3\mathbf{p} f(\mathbf{x}, \mathbf{p}, t) \quad (2.47)$$

$$\mathbf{u}(\mathbf{x}, t) = \frac{1}{\rho(\mathbf{x}, t)} \int_{\mathbf{p}} d^3\mathbf{p} \frac{\mathbf{p}}{ma} f(\mathbf{x}, \mathbf{p}, t). \quad (2.48)$$

Le premier moment donne l'équation de continuité :

$$\frac{\partial \delta}{\partial t}(\mathbf{x}, t) + \frac{1}{a} \text{div}_{\mathbf{x}} [(1 + \delta(\mathbf{x}, t)) \mathbf{u}(\mathbf{x}, t)] = 0. \quad (2.49)$$

<sup>2</sup>cela n'est vrai que pour les particules de matière noire

Le deuxième moment donne l'équation d'Euler :

$$\frac{\partial \mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}(\mathbf{u} \cdot \mathbf{grad}_{\mathbf{x}})\mathbf{u} = -\frac{1}{a}\mathbf{grad}_{\mathbf{x}}\Phi(\mathbf{x}, t) - \frac{1}{\rho}\frac{\partial}{\partial x_j}(\rho\sigma_{i,j}^2(\mathbf{x}, t)), \quad (2.50)$$

avec  $\sigma_{i,j}^2$  la pression dynamique du système qui est reliée aux moments d'ordre supérieur de l'équation de Boltzmann. Nous négligerons ce terme dans la suite.<sup>3</sup>

Nous allons néanmoins nous intéresser principalement à la première des approximations possibles : la théorie *linéaire* Eulérienne de croissance des fluctuations.

### 2.3.2 Théorie linéaire

Cette théorie consiste à supposer que les fluctuations et les vitesses sont “petites”. Quantitativement, on va imposer que

$$|\delta(\mathbf{x}, t)| \ll 1 \quad (2.51)$$

$$\frac{1}{a} \left| \frac{\partial u_i}{\partial x_j}(\mathbf{x}, t) \right| \ll H \quad (2.52)$$

avec  $u_i$  et  $x_j$  les composantes des vecteurs  $\mathbf{u}$  et  $\mathbf{x}$  respectivement. En linéarisant le système d'équation (2.49), (2.50), on obtient

$$\frac{\partial \delta}{\partial t}(\mathbf{x}, t) + \frac{1}{a} \text{div}_{\mathbf{x}} \mathbf{u} = 0, \quad (2.53)$$

$$\frac{\partial^2 \delta}{\partial t^2}(\mathbf{x}, t) + 2H \frac{\partial \delta}{\partial t}(\mathbf{x}, t) - 4\pi G \bar{\rho}(t) \delta(\mathbf{x}, t) = 0. \quad (2.54)$$

Nous allons d'abord nous intéresser à la deuxième équation. Celle-ci nous permet de séparer, dans  $\delta(\mathbf{x}, t)$ , la partie spatiale et temporelle. Comme (2.54) est une équation linéaire du second ordre, elle peut être séparée en deux solutions indépendantes, donc nous pouvons écrire d'une manière tout-à-fait générale :

$$\delta(\mathbf{x}, t) = D_+(t)\delta_+(\mathbf{x}) + D_-(t)\delta_-(\mathbf{x})\theta(\mathbf{x}, t) = \frac{1}{a\dot{D}}\mathbf{u}(\mathbf{x}, t) \quad (2.55)$$

avec  $\delta_+$  et  $\delta_-$  représentant le champ de densité initial. Les fonctions  $D_+(t)$  et  $D_-(t)$  sont appelées facteur de croissance linéaire. Plus spécifiquement,  $D_+(t)$  va décrire l'évolution du champ de densité dont les fluctuations croissent avec le temps, alors que  $D_-(t)$  concerne l'évolution pour laquelle les fluctuations décroissent. En introduisant l'égalité (2.55) dans (2.54), on obtient l'équation différentielle satisfaite par  $D_{\pm}(t)$

$$\ddot{D}_{\pm} + 2H\dot{D}_{\pm} - 4\pi G\bar{\rho}(t)D_{\pm}(t) = 0, \quad (2.56)$$

soit aussi, avec  $\bar{\rho}(t) = \frac{3}{2}\Omega_m(t)H^2(t)$ ,

$$\ddot{D}_{\pm} + 2H\dot{D}_{\pm} - \frac{3}{2}H^2(t)\Omega_m(t)D_{\pm}(t) = 0. \quad (2.57)$$

Quelques solutions de cette équation pour différents modèles d'Univers ont été proposé par Bernardeau et al. (2002) :

<sup>3</sup>On peut aussi le considérer comme une vraie pression. A l'aide d'un développement de la pression par rapport à la densité, on peut montrer qu'il existe une échelle critique, appelé longueur de Jeans, au-dessous de laquelle la dynamique est dominée par ce terme de pression. La discussion de la section suivante, basée sur cette approximation, n'est donc pas valide à ces échelles.



Modèle	Solution
$\Omega_m = 1, \Omega_\Lambda = 0$	$D_+(t) = a(t), D_-(t) = a^{-3/2}(t), \beta = 1$
$\Omega_m < 1, \Omega_\Lambda = 0$	<p>on pose <math>x = \frac{1}{\Omega_m(t)} - 1</math>, alors</p> $D_+ = 1 + \frac{3}{x} + 3\sqrt{\frac{1+x}{x^3}} \ln [\sqrt{1+x} - x]$ $D_- = \sqrt{\frac{1+x}{x^3}}$ <p>et <math>\beta</math> peut être approximé par <math>\beta \simeq \Omega_m^{3/5}</math>. Pour <math>\Omega_m \rightarrow 0</math>, alors <math>D_+ \rightarrow 1</math> et <math>D_0 \rightarrow \frac{1}{x}</math>, donc les perturbations cessent de grandir.</p>
Seulement matière et énergie sombre	<p>On ne peut plus résoudre analytiquement <math>D_+</math> (quoiqu'on le puisse pour <math>D_-</math>), mais on peut trouver une bonne approximation :</p> $D_+ \simeq \frac{5}{2} \frac{a\Omega_m}{\Omega_m^{4/7} - \Omega_\Lambda + (1 + \Omega_m/2)(1 + \Omega_\Lambda/70)},$ $D_- = \frac{H(a)}{a},$ $\beta \simeq \frac{1}{[1 - (\Omega_0 + \Omega_\Lambda^0 - 1)a + \Omega_\Lambda^0 a^3]^{0.6}},$ <p>où <math>\Omega_\Lambda^0 = \Omega_\Lambda(a = 1)</math>. Cette dernière égalité peut être grandement simplifiée dans le cas <math>\Omega_m + \Omega_\Lambda = 1</math> :</p> <div style="text-align: right;"> <math display="block">\boxed{\beta \simeq \Omega_m^{5/9}} \quad (2.58)</math> </div>

Nous ne nous intéresserons qu'au mode croissant  $D_+$  par la suite, que nous noterons en général  $D$ , celui-ci devenant très rapidement dominant dans la description de la dynamique de l'Univers aux grandes échelles.

À présent que nous connaissons bien l'évolution temporelle du champ des fluctuations, nous pouvons nous intéresser aux propriétés que nous donne la première équation (2.53). On peut exprimer la vitesse réduite  $\boldsymbol{\theta}(\mathbf{x})$  comme

$$\boldsymbol{\theta}(\mathbf{x}, t) = \frac{1}{a\dot{D}} \mathbf{u}(\mathbf{x}, t) \quad (2.59)$$

Si on utilise l'égalité (2.55) dans (2.53), on obtient

$$\text{div}_{\mathbf{x}} \boldsymbol{\theta}(\mathbf{x}) = -\delta(\mathbf{x}). \quad (2.60)$$

Nous ne savons pas si ce champ est irrotationnel, pour cela il faut utiliser la version linéarisée, où on néglige le terme convection, de (2.50). Si on prend le rotationnel, noté  $\mathbf{w} = \text{rot}_{\mathbf{x}} \mathbf{u}$ , de cette égalité on trouve

$$\frac{\partial \mathbf{w}}{\partial t} = -H \mathbf{w} \quad (2.61)$$

ce qui correspond à dire que la vorticit   $\mathbf{w}$  d croit au moins exponentiellement avec le temps, car  $H > 0$    tout instant.  $\mathbf{w}$  devient donc rapidement "nul". On peut alors r  crire (2.60) sous forme

intégrale, en supposant maintenant le champ de vitesse irrotationnel,

$$\theta(\mathbf{x}) = \frac{1}{4\pi} \int_{\mathbf{y}} d^3\mathbf{y} \delta(\mathbf{y}) \frac{\mathbf{y} - \mathbf{x}}{\|\mathbf{y} - \mathbf{x}\|^3}. \quad (2.62)$$

Enfin, on peut réexprimer  $\mathbf{u}(\mathbf{x}, t)$  en fonction de cette intégrale et après quelques manipulations algébriques :<sup>4</sup>

$$\mathbf{u}(\mathbf{x}, t) = \frac{aH\beta}{4\pi} \int_{\mathbf{y}} d^3\mathbf{y} \delta(\mathbf{y}, t) \frac{\mathbf{y} - \mathbf{x}}{\|\mathbf{y} - \mathbf{x}\|^3} \quad (2.63)$$

avec  $\beta = d \log D / d \log a$ . Cette relation fait apparaître que champ de vitesse est proportionnel au champ de gravitation au même instant.

### 2.3.3 L'approche Lagrangienne : l'approximation de Zel'dovich

L'approximation de Zel'dovich (Zel'Dovich, 1970) est une approximation Lagrangienne du premier ordre à la dynamique non-collisionnelle des particules de matière noire. Nous supposons que nous connaissons l'application  $\mathbf{q} \rightarrow \mathbf{x}(\mathbf{q}, t)$  qui représente la position à un instant  $t$  d'une particule de matière noire qui était originellement, *i.e.* à l'instant  $t = 0$ , à la position  $\mathbf{q}$ . Sa position peut être décrite par l'équation suivante :

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + \Psi(\mathbf{q}, t) \stackrel{\text{ZA}}{\simeq} \mathbf{q} + \tilde{D}(t) \Psi_0(\mathbf{q}), \quad (2.64)$$

avec  $\tilde{D}(t) = 0$  le facteur de croissance du champ de déplacement,  $\Psi_0(\mathbf{q}) = \psi(\mathbf{q}, t_{\text{maintenant}})$ . La dernière égalité correspond à l'approximation de Zel'dovich, qui considère que le temps et l'espace sont des variables séparables. La fonction  $\tilde{D}(t)$  sera déterminée par la suite.

Pour  $\tilde{D}(t) \ll 1$ , l'équation (2.64) est liée au champ de densité  $\rho(\mathbf{x}, t)$  par l'intermédiaire de la conservation de la masse entre l'instant  $t = 0$ , où l'Univers est supposé homogène, et l'instant  $t$  :

$$\rho_0 d^3 \mathbf{q} = \rho(\mathbf{x}, t) d^3 \mathbf{x} \quad (2.65)$$

$$= \rho(\mathbf{x}, t) \det \left( \delta_{i,j} + \frac{\partial \Psi_i}{\partial q_j} \right) d^3 \mathbf{q} \quad (2.66)$$

Cette relation nous mène à l'égalité suivante, en supposant  $\frac{\partial \Psi_i}{\partial q_j} \ll 1$  :

$$\frac{\rho(\mathbf{x}, t)}{\rho_0} \simeq 1 - \text{div}_{\mathbf{q}} (\Psi(\mathbf{q}, t))|_{\mathbf{q}(\mathbf{x})} \quad (2.67)$$

Le champ de gravité  $\mathbf{g}(\mathbf{x}) = \mathbf{grad}_{\mathbf{x}} \Phi$  est donné par

$$\text{div}_{\mathbf{x}} \mathbf{g} = -4\pi G \rho_0 \left( \frac{\rho(x, t)}{\rho_0} - 1 \right) \quad (2.68)$$

avec  $G$  la constante de gravitation universelle. Ceci montre que

$$\Psi_0(\mathbf{q}) = \frac{1}{\alpha D(t)} (\mathbf{grad}_{\mathbf{x}} \Phi(\mathbf{x}, t))_{\mathbf{q}} \stackrel{t \rightarrow 0}{=} \mathbf{grad}_{\mathbf{q}} \Xi(\mathbf{q}), \quad (2.69)$$

car pour  $t \rightarrow 0$ , les coordonnées  $\mathbf{x}$  et  $\mathbf{q}$  sont identiques. Maintenant en écrivant

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{grad}_{\mathbf{q}} \xi(\mathbf{q}, t), \quad (2.70)$$

on peut trouver que le potentiel du champ de déplacement est

$$\xi(\mathbf{q}, t) = \frac{\|\mathbf{q}\|_2^2}{2} + \tilde{D}(t) \Xi(\mathbf{q}). \quad (2.71)$$

<sup>4</sup>Noter que maintenant  $\delta$  dépend du temps.

En utilisant les équations (2.67) et (2.64), on peut relier la densité eulérienne  $\delta$  au champ de déplacement :

$$\delta(\mathbf{x}, t) = -\tilde{D}(t) \operatorname{div}_{\mathbf{q}} (\Psi_0(\mathbf{q}))|_{\mathbf{q}(\mathbf{x})} = \tilde{D}(t) \delta(\mathbf{x}). \quad (2.72)$$

Par identification avec l'égalité (2.55), nous pouvons à présent identifier  $\tilde{D}(t)$  à  $D(t)$  le facteur de croissance des fluctuations en théorie Eulérienne. Le champ de vitesse comobile prédit est donc

$$\mathbf{v}(\mathbf{x}, t) \equiv \frac{\partial \mathbf{x}(\mathbf{q}, t)}{\partial t} = \dot{D}(t) \Psi_0(\mathbf{q}(\mathbf{x})) = \frac{1}{a} \mathbf{u}(\mathbf{x}, t). \quad (2.73)$$

L'introduction de la solution du facteur de croissance  $D(t)$  dans l'équation (2.71) nous permet d'obtenir la position d'une particule de matière noire à tout instant  $t$ , pourvu que l'approximation de Zel'dovich reste valide à cet instant et à la position considérée. On peut ré-écrire l'équation (2.73) en fonction du facteur d'échelle  $a$  à la place de la variable temporelle  $t$ . Pour cela on remplace  $\dot{D}(t)$  :

$$\dot{D}(t) = \dot{a} \frac{dD}{da} = HD \frac{d \log D}{d \log a} = HD\beta, \quad (2.74)$$

ce qui nous donne

$$\mathbf{v}(\mathbf{x}, a) = H\beta D(a) \Psi_0(\mathbf{q}(\mathbf{x})) = H\beta (\mathbf{x}(\mathbf{q}, t_{\text{maintenant}}) - \mathbf{q}). \quad (2.75)$$

On a donc une relation simple entre le déplacement total d'une particule de sa position à l'instant d'origine à celle à l'instant présent et sa vitesse actuelle. On peut noter que les trajectoires, dans l'approximation de Zel'dovich, sont rectilignes mais que le mouvement n'est pas uniforme. Il ne s'agit donc pas d'une approximation purement inertielle. Par ailleurs, nous verrons par la suite que l'approximation des trajectoire rectilignes est non seulement bonne mais aussi robuste.

## 2.4 Intérêt et obtention des champs de vitesse

### 2.4.1 Intérêt des champs de vitesse

Comme indiqué dans l'introduction, le champ de vitesses est un indicateur direct de la vraie distribution locale de matière noire. On peut le voir directement grâce à l'équation (2.63). Cette relation exprime que, dans la limite de la théorie linéaire, le champ de vitesse est proportionnel au champ de gravitation dû à la distribution de *masse*. Un autre intérêt du champ de vitesse apparaît si on réexprime (2.63) dans l'espace des modes de Fourier. Pour cela on va supposer que le champ de vitesse est potentiel (cette approximation est motivée par la décroissance rapide de la vorticit  avec le temps). On a donc  $\theta = -\operatorname{grad}_{\mathbf{x}} \alpha$ . En injectant dans l'équation (2.60), on obtient

$$\Delta_{\mathbf{x}} \alpha = -\delta(\mathbf{x}) \quad (2.76)$$

soit, après transformation de Fourier,

$$k^2 \mathcal{F}(\alpha) = \mathcal{F}(\delta) \quad (2.77)$$

Par conséquent, (2.63) peut  tre r - crite :

$$\hat{\mathbf{u}}(\mathbf{k}, t) = \mathcal{F}(\mathbf{u}) = -\frac{i\mathbf{k}}{k^2} \hat{\delta}(\mathbf{k}) \quad (2.78)$$

Comparativement au champ de densit , le champ de vitesse est donc beaucoup plus sensible aux grandes  chelles. En effet, dans l' quation ci-dessus, les modes de Fourier du champ de vitesse correspondent aux modes du champ de densit  pond r  par un facteur  $1/k$  qui est grand aux grandes  chelles (car  $k$  est alors petit). Cela signifie que le champ de vitesse sera  galement moins affect  par la dynamique non-lin aire que le champ de densit . La figure 2.4 pr sente un sch ma sur lequel sont positionn es les diff rentes exp riences en cours ou d j  r alis es, associ es aux r gions du spectre de puissance dont elles constituent une sonde. On esp re sonder les r gions indiqu es par la fl che rouge pour le champ de vitesse.

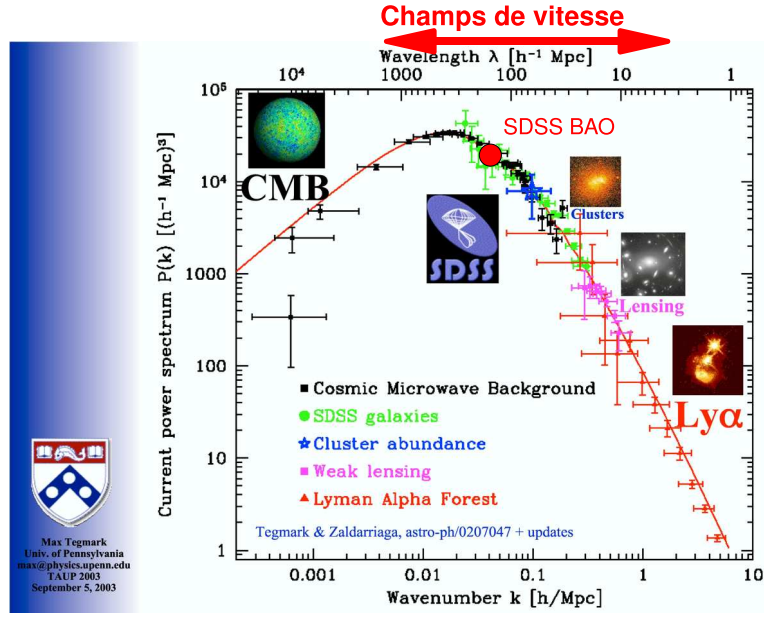


FIG. 2.4: *Champ de vitesse et spectre de puissance* – Illustration des régions du spectre de puissance auxquelles sont sensibles les différentes expériences en cours. Le régime des champs de vitesse est indiqué au-dessus. On est intrinsèquement limité à  $5 h^{-1}\text{Mpc}$  pour la méthode que l'on va exposer dans la suite et la méthode ne présente à priori pas de limite supérieure si ce n'est la profondeur des catalogues que l'on peut obtenir (pour l'instant  $\sim 1 h^{-1}\text{Gpc}$ ).

#### 2.4.2 Les distances dans l'Univers / Mesurer les vitesses

Malheureusement pour nous, il est difficile de mesurer les vitesses propres des galaxies qui sont situées trop loin pour que les méthodes astrométriques standards puissent être utilisées (par exemple la méthode des parallaxes). On doit donc utiliser des *indicateurs* de distance qui ont été obtenus et calibrés empiriquement à partir d'objets de l'Univers extrêmement proche. Ces indicateurs sont basés typiquement sur une caractéristique physique observée qui ne dépend pas de la distance à l'objet.

Plusieurs indicateurs ont été construits. On peut en citer quelques-uns dont :<sup>5</sup>

1. Les céphéides : ce sont des étoiles variables dont l'atmosphère externe pulse avec une période de 2 à 100 jours. Elles sont particulièrement abondantes dans les galaxies du Volume local ( $\sim 10 \text{ Mpc/h}$ ). On connaît très bien théoriquement le mécanisme de la pulsation et on a établi empiriquement que la période de la pulsation, qui ne dépend pas de la distance, est liée à la luminosité intrinsèque de l'étoile. Il existe une dispersion de l'ordre de 20% en luminosité dans cette relation, ce qui correspond à une erreur de 10% sur la distance. On présente, figure 2.5, le diagramme de Hubble obtenu par le HST Key Project (Freedman et al., 2001a). Malheureusement, cette méthode, assez précise, n'est pas utilisable pour le moment au-delà de  $\sim 20 h^{-1}\text{Mpc}$ , distance qui ne permet plus la détection des céphéides. Nous devons donc utiliser des objets plus lumineux.
2. Supernovae de type Ia : Les supernovae de type Ia (SN Ia) sont des indicateurs de distance potentiellement très performants pour la cosmologie observationnelle moderne. La relation établie entre le temps caractéristique de leur évolution et leur luminosité maximale est utilisée comme indicateur. L'incertitude sur la luminosité est de l'ordre de 12%, ce qui nous amène à une incertitude de 6% sur la distance. Comme leur luminosité est de l'ordre de grandeur d'une galaxie de taille moyenne, cela rend les SN Ia les indicateurs de distance les plus précis et les plus pro-

<sup>5</sup>Nous conseillons au lecteur avide de détails de se reporter à la revue Freedman (2000).

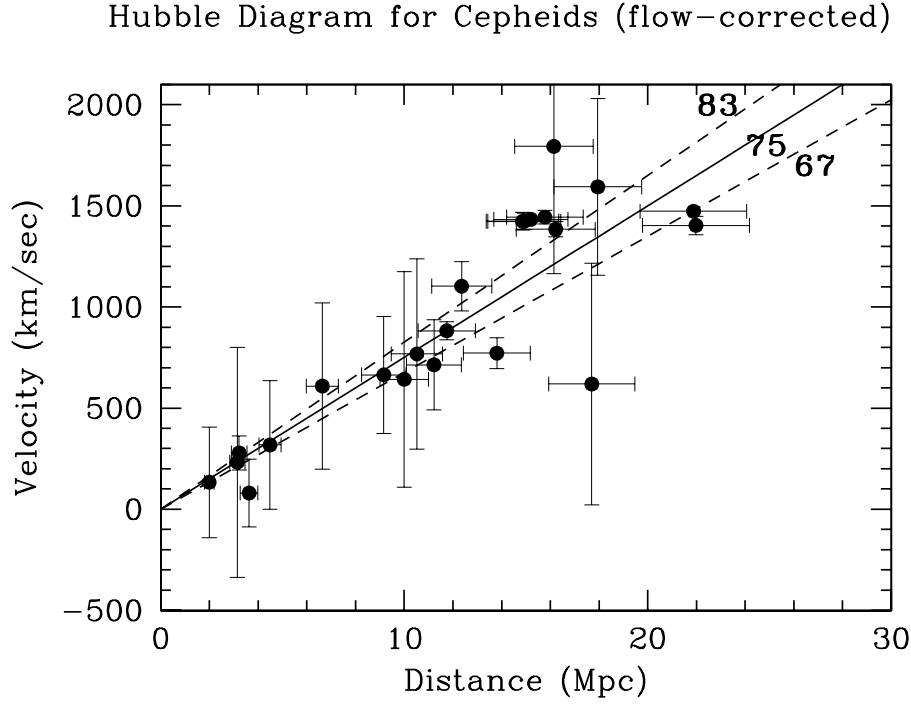


FIG. 2.5: *Les distances dans l'Univers* – Déduction de la constante de Hubble en utilisant les distances mesurées des céphéides comparées à leur décalage vers le rouge. Cette figure a été extraite de l'article du Hubble Space Telescope Key Project (Freedman et al., 2001a). Leur meilleure estimation est de l'ordre de  $H_0 \simeq 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

fonds de la cosmologie. Néanmoins, le mécanisme de l'explosion n'a pas encore été bien expliqué observationnellement et théoriquement.

3. La relation de Tully-Fisher : cette relation lie empiriquement la luminosité totale d'une galaxie à sa vitesse de rotation (voir la description plus détaillée à la section 4.3.1). Intuitivement, on peut comprendre que plus une galaxie est lumineuse, plus elle est massive et donc plus elle doit tourner rapidement pour compenser l'attraction gravitationnelle. On peut mesurer les vitesses de rotation grâce à l'effet Doppler en mesurant l'élargissement de la raie à 21cm du gaz HI dans le domaine radio. L'erreur typique sur la luminosité est de l'ordre de 30%, ce qui conduit à une estimation de l'erreur 15% de la distance. La principale source d'erreur dans la méthode est l'effet de projection quand la galaxie n'est pas observée exactement de profil. Comme il est nécessaire d'avoir une galaxie en rotation, cette relation n'est applicable que pour les galaxies spirales.
4. La plan fondamental : Il s'agit d'une relation analogue à celle de Tully-Fisher, mais qui concerne les galaxies elliptiques seules. On relie cette fois la luminosité intrinsèque à la dispersion des étoiles qui s'y trouvent et à la brillance de surface. On a typiquement une relation  $L \propto I_0^{-0.7} \sigma^3$ . La précision de cette relation est de l'ordre de 35-40%, ce qui donne une précision sur la distance de 20%.
5. Fluctuation de la brillance de surface : la résolution des étoiles par une caméra CCD dépend de la distance. Dans chaque pixel de la CCD, un nombre donné d'étoiles contribue à la luminosité. Les fluctuations de Poisson d'un pixel à un autre dépendent de la distance de la galaxie. On a constaté que ces fluctuations dépendaient très fortement de la couleur des étoiles. Cette méthode a été développée par Tonry & Schneider (1988) et a une précision de l'ordre de 8%. Elle peut être appliquée jusqu'à des décalages spectraux  $z \sim 0.02$  (soit  $\sim 60 h^{-1} \text{ Mpc}$ ).
6. Pointe de la branche des géantes rouges (Tip of the Red Giant Branch, TRGB Lee et al., 1993; Karachentsev et al., 2004, 2006). Cette méthode s'appuie sur la luminosité standard des étoiles

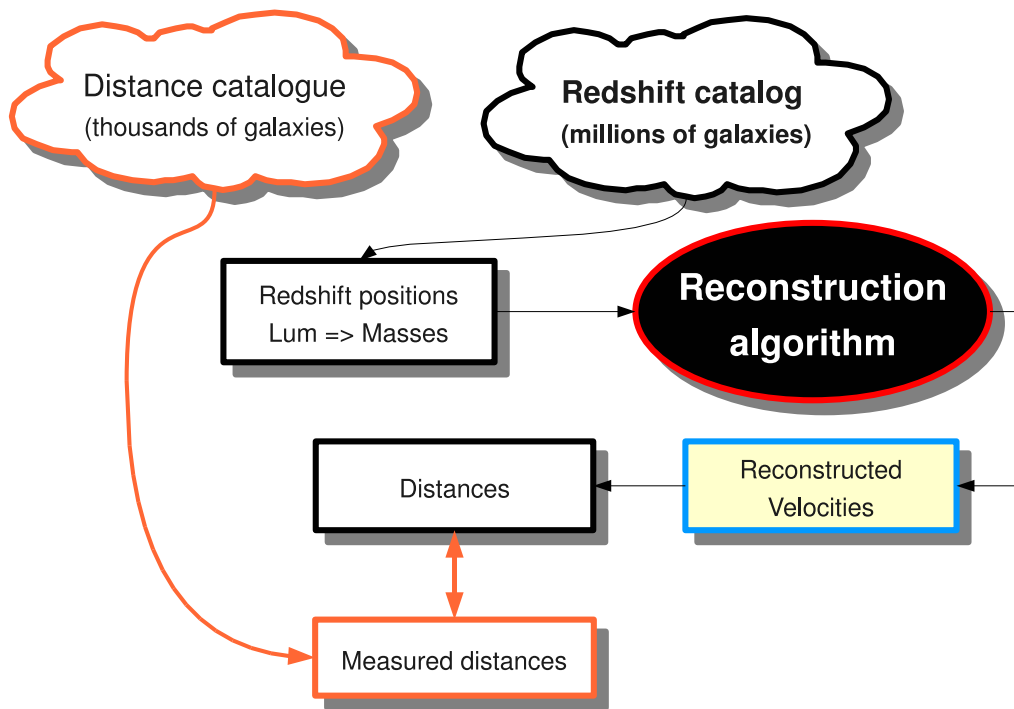


FIG. 2.6: Diagramme illustrant le processus général de reconstruction du champ de vitesse

géantes rouges lors de leur évolution dans le diagramme couleur-magnitude. Un instant de leur vie est décrit par un point prenant la forme d'un doigt dans ce diagramme, qui est donc aisément reconnaissable et mesurable. Elle présente une précision d'environ 20% sur la luminosité, soit 10% sur la distance.

### 2.4.3 Les méthodes de reconstruction

Dans tous les cas, les méthodes ci-dessus ne donnent accès qu'à la composante radiale des vitesses de galaxie et ceci avec un bruit qui peut être rédhibitoire. On souhaiterait avoir accès aux trois composantes du champ de vitesse. Pour cela, on peut utiliser des méthodes de reconstruction de ce champ. Il existe pour cela deux écoles :

- une extrapolation du champ utilisé en conjonction avec un lissage (méthode ci-après POTENT)
- une prédiction du champ de vitesse obtenu grâce à la distribution de masse que l'on peut comparer a posteriori avec les distances observées de chaque galaxies (méthode Least-Action et Monge-Ampère-Kantorovitch).

De manière générale, la procédure dans le cas des méthode de reconstruction lagrangienne (deuxième item ci-dessus) peut se décrire à l'aide du diagramme représenté figure 2.6. A partir d'un catalogue de positions des galaxies avec leur redshift (actuellement on peut compter quelques millions de galaxies), on synthétise une distribution de masse que l'on obtient à l'aide d'une relation entre la luminosité, la masse et la distribution des objets en espace redshift. Cette distribution est introduite dans une boîte noire (l'algorithme de reconstruction), qui produit des vitesses reconstruites. On peut alors calculer les vraies distances des galaxies du catalogue d'entrée. Ces distances peuvent à leur tour être comparées celles proposées par un second catalogue de galaxies dont la valeur des distances a pu être renseignée par une méthode alternative (au maximum quelques milliers de galaxies actuellement). Le résultat apporté par ce processus de comparaison permet de valider le procédé de fabrication de la distribution des masses après un ajustement éventuel des hypothèses retenues. Une fois que la procédure est validée, on peut se servir du champ de vitesse reconstruit pour extraire des informations sur les paramètres cosmologiques.

### 2.4.4 La méthode POTENT

Une première méthode de reconstruction du champ de vitesse à partir d'observations a été proposée par Bertschinger & Dekel (1989) et de manière plus détaillée par Dekel et al. (1990). Il s'agit d'une méthode tentant de reconstruire le champ vectoriel de vitesse particulière à partir d'un échantillonnage de la composante radiale de ce champ. Cette méthode a été dénommée POTENT par les auteurs et nous garderons cette appellation ici. POTENT s'appuie sur l'hypothèse que le champ de vitesse  $\mathbf{v}(\mathbf{x}, t)$  tracé par les galaxies est irrotationnel en coordonnée comobile lagrangienne  $\mathbf{q}$ , c'est-à-dire :

$$\mathbf{rot}_{\mathbf{q}} \mathbf{u} = 0 . \quad (2.79)$$

Cette propriété non triviale est motivée, aux grandes échelles, par la théorie linéaire qui stipule que la vorticit  du champ de vitesse s'annule rapidement (cf. Peebles, 1980, ou section 2.3.2). Comme le champ de gravit  est irrotationnel, il doit en  tre de m me pour le champ de vitesse. On peut alors poser qu'il existe un champ scalaire  $\Phi$  dont d rive le champ de vitesse :

$$\mathbf{u}(\mathbf{q}, a) = \mathbf{grad}_{\mathbf{q}} \Phi(\mathbf{q}, a) \quad (2.80)$$

En supposant que l'on connaisse le champ scalaire

$$u_r(\mathbf{q}, a = 1) = \mathbf{u}(\mathbf{q}, a = 1) \cdot \hat{\mathbf{q}} = \frac{\partial \Phi}{\partial q_r} , \quad (2.81)$$

avec  $q_r = \|\mathbf{q}\|$ , on peut obtenir le potentiel scalaire  $\Phi$  par int gration

$$\Phi(\mathbf{q}(q_r, \theta, \phi), a) = \int_0^{q_r} u_r(\mathbf{s}(Q_r, \theta, \phi)) dQ_r . \quad (2.82)$$

o   $q_r$  (respectivement  $Q_r$ ),  $\theta$ ,  $\phi$  sont les coordonn es sph riques de  $\mathbf{q}$  (respectivement  $\mathbf{s}$ ).

La figure 2.7 montre une reconstruction du champ de vitesse par la m thode POTENT appliqu e   un champ de vitesse radiale calcul e par une simulation de mati re noire. Evidemment, plusieurs probl mes fondamentaux se posent dans la pratique :

- nous n'avons pas acc s aux coordonn es lagrangiennes du traceur de vitesse dans le catalogue (une galaxie ou un groupe de galaxies),
- nous n'avons acc s qu'  des traceurs et non au champ de vitesse lui-m me.

Parmi ces deux limitations, il s'agit en fait du premier point qui pose le plus de probl me   la reconstruction.

Dekel et al. (1990) proposent des solutions afin de tenir compte des probl mes ci-dessus. Notamment, les distortions de redshift par une m thode it rative sur le champ de vitesse  $\mathbf{u}(\mathbf{q})$ . Cependant, la convergence de cette m thode n'est jamais garantie et est m me assez fragile quand elle appliqu e sur des donn es bruit es. Il faut finalement pr ciser que cette m thode utilise intrins quement la th orie perturbative eul rienne et donc qu'elle aura des difficult s   fonctionner dans les r gimes non-lin aires.

### 2.4.5 La m thode de moindre action

Une approche radicalement diff rente de POTENT est celle de la m thode de moindre action, en anglais Least-Action (Peebles, 1989). Il s'agit en fait d'une m thode   gravit  n gative, o  on essaye de repousser les galaxies d'un catalogue   leur emplacement comobile d'origine. Mis   part le fait que les galaxies (ou les groupes de galaxies) devraient  tre form es progressivement,<sup>6</sup> il n'y aurait, en principe, aucun probl me avec cette m thode si ce n'est qu'on ne conna t justement pas la vitesse actuelle des objets. Cette absence de contraintes entra nent que le nombre de solutions plausibles qui sont trouv es par la m thode cro t tr s fortement avec le volume et/ou le nombre d'objets consid r s. La m thode

<sup>6</sup>Une premi re tentative pour tenir compte de ce probl me est propos e par Phelps et al. (2006). Grossi rement d crite, la m thode consiste   faire accr ter progressivement de la masse aux objets du catalogue alors qu'ils se d placent vers leur position actuelle.

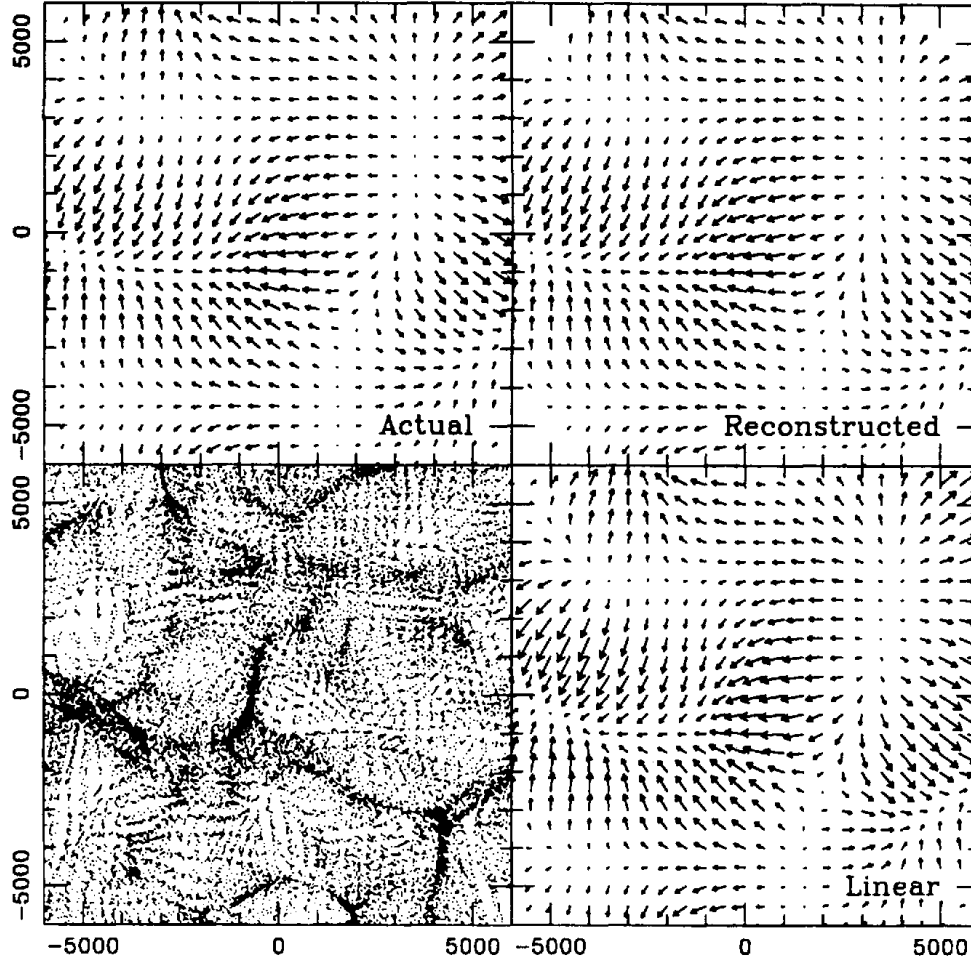


FIG. 2.7: Reconstruction POTENT, figure extraite de Dekel et al. (1990). Cette figure représente le champ de vitesse bidimensionnel dans un plan d’une simulation à  $N$  corps de matière noire. Les unités des axes sont exprimés en km/s et l’espacement entre les flèches est de 500 km/s. Le panneau nommé “Actual” montre le champ de vitesse, une fois lissé, de la simulation. Le panneau “Reconstructed” montre le champ de vitesse reconstruit par la méthode POTENT à partir de seulement la composante radiale du champ de vitesse “Actual”. Le panneau marqué “Linear” utilise l’approche perturbative Lagrangienne du premier ordre (§ 2.3.3) pour calculer le champ de vitesse final de la simulation à partir des conditions initiales.



essaye de contourner ce problème en fixant les conditions de bord des équations du mouvement au lieu de fixer l'espace des phases des conditions finales.

Nous allons décrire ici brièvement les principes de la méthode Numerical-Action-Model (Nusser & Branchini, 2000). Le formalisme qui suit suppose que l'on peut négliger les effets de marée sur le volume  $V$  d'Univers considéré.<sup>7</sup> Nous allons utiliser le taux de variation  $D(t)$  du mode croissant de la théorie linéaire (Peebles, 1980) pour représenter le temps. En posant que  $\mathbf{x}_i$  est la position d'une galaxie  $i$  et que  $\boldsymbol{\theta}_i = \frac{d\mathbf{x}_i}{dD}$ , la galaxie  $i$  obéit à l'équation suivante (Nusser & Colberg, 1998; Nusser & Branchini, 2000) :

$$\frac{d\boldsymbol{\theta}_i}{dD} + \frac{3}{2} \frac{\Omega_m}{f(\Omega)^2} \frac{\boldsymbol{\theta}_i}{D} = \frac{3}{2} \frac{1}{D^2} \frac{\Omega_m}{f^2(\Omega_m)} \mathbf{g}(\mathbf{x}_i), \quad (2.83)$$

avec  $f(\Omega_m) = \frac{d \log D}{d \log a} \simeq \Omega_m^{5/9}$  et  $\mathbf{g}(\mathbf{x})$  le champ de gravitation évalué au point  $\mathbf{x}$ . Cette équation est obtenue en exprimant les équation de Newton en coordonnées comobile (§ 2.2.6). En fait, on constate que le membre de gauche peut devenir très simple si on fait l'approximation  $f^2(\Omega_m) \simeq \Omega_m^{1.1} \simeq \Omega_m$ . Différentes approximations de ce genre ont été étudiée par Nusser & Colberg (1998), ce qui les a amenés à utiliser la forme approchée suivante des équations du mouvement :

$$\frac{d\boldsymbol{\theta}_i}{dD} + \frac{3}{2} \frac{\boldsymbol{\theta}_i}{D} = \frac{1}{D^{3/2}} \frac{d(D^{3/2}\boldsymbol{\theta}_i)}{dD} = \frac{3}{2} \frac{1}{D^2} \frac{\Omega_m(D)}{f^2(\Omega_m(D))} \mathbf{g}(\mathbf{x}_i). \quad (2.84)$$

Après quelques transformations il est possible de réexprimer cette équation d'évolution sous la forme d'une intégrale d'action :

$$S = \int_0^1 dD \sum_i \left[ \frac{1}{2} D^{3/2} \boldsymbol{\theta}_i^2 + \frac{3}{2} \frac{1}{D^{1/2}} \frac{\Omega_m}{f(\Omega_m)^2} \left( \frac{1}{4\pi\bar{n}} \sum_{j<i} \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|} + \frac{\mathbf{x}_i^2}{6} \right) \right]. \quad (2.85)$$

La méthode de moindre action ne va pas chercher à minimiser  $S$  sur absolument *tous* les chemins possibles. Elle va se limiter à étudier un sous-ensemble de trajectoires décrit par une base de fonction  $\{q_n(D)\}$ .<sup>8</sup> On exprimera donc les  $\mathbf{x}_i$  et les  $\boldsymbol{\theta}_i$  sur cette base :

$$\mathbf{x}_i(D) = \mathbf{x}_{i,0} + \sum_{j=1}^N q_j(D) \mathbf{C}_i^j, \quad (2.86)$$

$$\boldsymbol{\theta}_i(D) = \sum_{j=1}^N p_j(D) \mathbf{C}_i^j, \quad (2.87)$$

avec  $p_j(D) = \frac{dq_j}{dD}$ . A présent, nous pouvons explicitement calculer la minimisation de l'action  $S$  par rapport aux  $\mathbf{C}_i^j$  :

$$\frac{\partial S}{\partial \mathbf{C}_i^j} = \int_0^1 dD \left( D^{3/2} p_j(D) \boldsymbol{\theta}_i + \frac{3}{2} \frac{q_j}{D^{1/2}} \frac{\Omega_m}{f^2(\Omega_m)} \mathbf{g}(\mathbf{x}_i) \right) = 0 \quad (2.88)$$

En intégrant par partie cette équation, on trouve aisément :

$$\begin{aligned} \left[ D^{3/2} q_j(D) \boldsymbol{\theta}_i(D) \right]_{D=1} - \lim_{D \rightarrow 0} \left[ D^{3/2} q_j(D) \boldsymbol{\theta}_i(D) \right] \\ - \int_0^1 dD D^{3/2} q_j(D) \left( \frac{d\boldsymbol{\theta}_i}{dD} + \frac{3}{2D} \boldsymbol{\theta}_i - \frac{3}{2} \frac{1}{D^2} \frac{\Omega_m}{f^2(\Omega_m)} \mathbf{g}(\mathbf{x}) \right) = 0 \end{aligned} \quad (2.89)$$

<sup>7</sup>Ceci est faux en pratique mais permet de simplifier le formalisme utilis pour les besoins de la présente description.

<sup>8</sup>On peut noter l'idée d'approche perturbative Lagrangienne ici.

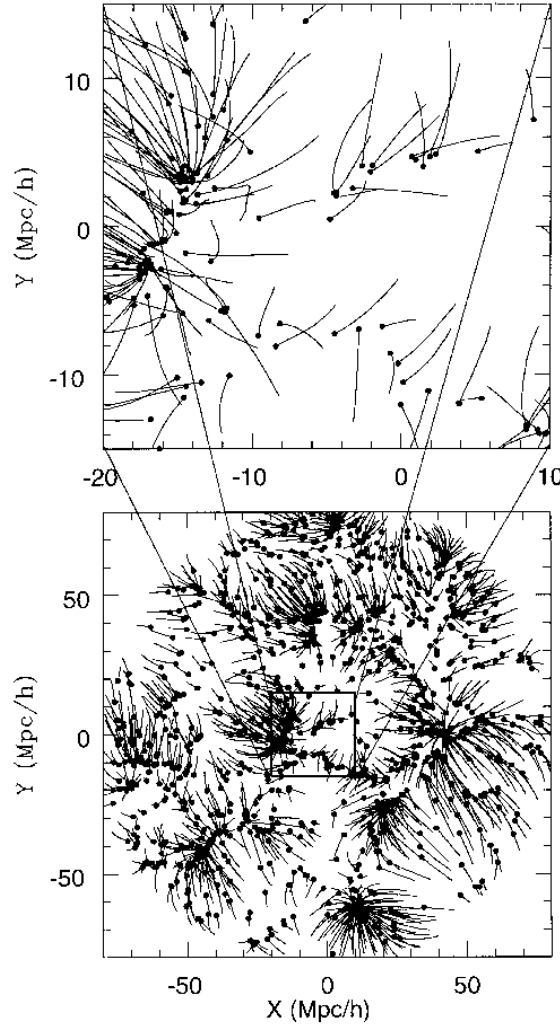


FIG. 2.8: Un exemple d'orbites reconstruites par la méthode Fast-Action-Method (figure extraite de Nusser & Branchini, 2000). Les points noir représentent les positions actuelles pour une sélection aléatoire de particules d'une simulation à  $N$  corps, contenues dans une tranche d'épaisseur  $10 h^{-1}\text{Mpc}$ . Les lignes solides représentent leur orbites projetées. Le panneau du haut est une version agrandie de la région sélectionnée en bas.

Comme l'expression sous le signe intégrale est obligatoirement nulle si les particules sont évoluées de manière à satisfaire l'équation (2.84), il ne reste qu'à annuler les deux premiers termes. Nous imposons alors les deux équations de contrainte suivantes sur les  $q_n(D)$  afin de respecter le principe de moindre action :

$$q_n(1) = 0 \quad \text{et} \quad \forall i, \lim_{D \rightarrow 0} D^{3/2} q_n(D) \theta_i(D) = 0 \quad (2.90)$$

Beaucoup de fonctions  $q_n(D)$  satisfont ces deux conditions. Nusser & Branchini (2000) proposent d'utiliser une base de fonctions orthogonales composées des monômes  $1, D, D^2, \dots$ . Il reste maintenant à minimiser l'action (2.85) par rapport aux  $C_i^j$ , les conditions de bord étant automatiquement vérifiées par notre choix de base de fonction  $p_n(D)$ . Nusser & Branchini (2000) utilisent une méthode de gradient conjugué (Press et al., 1992). Un exemple de reconstruction d'orbites obtenus par cette méthode est présenté sur la figure 2.4.5. En pratique, la qualité des vitesses reconstruites est très bonne par cette méthode. Néanmoins ses principaux inconvénients que l'on a décrit plus haut subsistent.

### 2.4.6 La reconstruction de Monge-Ampère-Kantorovitch (MAK)

Dans la plupart des modèles actuels, les particules de matière noire sont supposées évoluer sans collisions. Tout algorithme qui aurait pour but de prédire les vitesses particulières des galaxies devrait alors se baser sur les équations de Boltzmann. Mais si on se limite à étudier des régions dans l'Univers qui ne sont pas encore rentrées dans le régime des équations de Boltzmann,<sup>9</sup> il est possible de simplifier le problème en utilisant une description fluide de la matière noire.

On peut se fonder sur les résultats obtenus dans le cadre de l'approximation de Zel'dovich de la section 2.3.3. On peut constater que le potentiel  $\xi(\mathbf{q}, t)$  dans l'équation (2.71) est une fonction convexe. En effet, si on calcule la courbure  $\mathcal{C}_{i,j}$  de  $\xi(\mathbf{q}, t)$  :

$$\mathcal{C}_{i,j} = \frac{\partial^2 \xi}{\partial q_i \partial q_j} = \delta_{i,j} + \bar{D}(t) \frac{\partial \Xi}{\partial q_i \partial q_j}, \quad (2.91)$$

on constate que la courbure aura des valeurs propres strictement positives pourvu que  $\bar{D}(t)$  soit suffisamment petit. C'est une condition suffisante pour affirmer que  $\xi(\mathbf{q}, t)$  est alors une fonction convexe.  $\Psi_0(\mathbf{q})$  dérive donc d'un potentiel convexe pour un facteur de croissance  $D(t)$  suffisamment petit.

A présent nous allons renverser le raisonnement : nous *supposons* que l'application décrivant le changement de coordonnées Lagrangien-Eulérien  $\mathbf{q} \rightarrow \mathbf{x}(\mathbf{q}, t)$  dérive d'un potentiel convexe (hypothèse H) [comme dans l'équation (2.71)]. Cette hypothèse est *motivée* par l'efficacité de l'approximation de Zel'dovich aux grandes échelles (typiquement la taille d'un groupe de quelques galaxies, cf. chapitre 3). Aux petites échelles, les régions où le champ de vitesse est multivalué sont supposées être confinées dans les amas de galaxies ou les filaments. Nous allons donc résoudre *exactement* un problème approché, qui correspondrait plutôt à la dynamique donnée par le modèle d'adhésion (Gurbatov et al., 1989; Shandarin & Zeldovich, 1989). L'application satisfaisant l'hypothèse (H) est *contrainte* par la conservation locale de la masse. En introduisant l'égalité (2.70) dans l'équation (2.66), on obtient un problème de Monge-Ampère :

$$\det \left( \frac{\partial^2 \xi}{\partial q_j \partial q_k} \right) = \frac{\rho_0}{\rho(\mathbf{x}(\mathbf{q}), t)} \quad (2.92)$$

avec  $\rho(\mathbf{x}, t)$  le champ de densité (de masse) à l'instant  $t$ ,  $\rho_0$  le champ de densité initial, supposé homogène.<sup>10</sup> Cependant, l'équation (2.92) peut être simplifiée grâce à l'hypothèse de *convexité*. En posant

$$\mathbf{q}(\mathbf{x}) = \text{grad}_{\mathbf{x}} \Theta(\mathbf{x}) \quad (2.93)$$

l'application inverse qui nous permet de revenir des coordonnées eulériennes aux coordonnées lagrangiennes, on sait que l'on peut relier  $\xi$ , fonction convexe, à  $\Theta$ , qui sera automatiquement convexe, par une transformée de Legendre :

$$\Theta(\mathbf{x}) = \mathbf{q}(\mathbf{x}, t) \cdot \mathbf{x} - \xi(\mathbf{q}(\mathbf{x}, t)) . \quad (2.94)$$

Avec ces coordonnées, (2.92) s'écrit :

$$\det \left( \frac{\partial^2 \Theta}{\partial x_j \partial x_k} \right) = \frac{\rho(\mathbf{x}, t)}{\rho_0}, \quad (2.95)$$

qui ressemble très fortement à une équation de Poisson. En effet dans le cas linéaire on a  $\rho \simeq \rho_0$  et on peut développer à l'ordre 1 le déterminant dans le membre de gauche, ce qui revient à prendre la trace de la matrice :

$$1 + \Delta_{\mathbf{x}} \Theta \simeq \frac{\rho(\mathbf{x}, t)}{\rho_0}. \quad (2.96)$$

<sup>9</sup>c'est-à-dire qu'il n'y a pas encore de multivaluation du champ de vitesse

<sup>10</sup>Cette contrainte est motivée par les observations du fond diffus cosmologique. Cependant, on peut aussi montrer que pour assurer l'existence d'une solution continue aux équations d'Euler-Poisson, on doit supposer que le champ de densité initiale est homogène

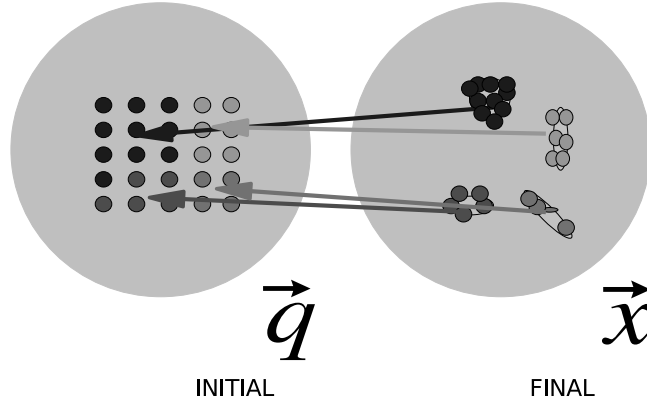


FIG. 2.9: *Reconstruction MAK* – L’application lagrangienne  $\mathbf{q}(\mathbf{x})$  qui lie le déplacement de la distribution de matière entre le rond de gauche et le rond de droite est calculé par l’intermédiaire de la méthode de MAK. La méthode de reconstruction MAK nous amènerait à assigner aux particules dans l’état “actuel”, représentées par la distribution dans le cercle de droite, les volumes convexes indiquées dans le diagramme de gauche.

Maintenant, il nous faut trouver un moyen de résoudre l’équation complète non-linéaire (2.95). Brenier (1987, 1991) a montré que résoudre l’équation (2.95) est équivalent à résoudre un problème de Kantorovitch :

$$I[\mathbf{q}(\mathbf{x})] = \int \rho(\mathbf{x}) |\mathbf{x} - \mathbf{q}(\mathbf{x})|^2 d^3 \mathbf{x} \quad (2.97)$$

par rapport à l’application Lagrangienne  $\mathbf{q}(\mathbf{x})$ , ou de manière équivalente, par rapport à  $\mathbf{x}(\mathbf{q})$ , grâce à l’hypothèse de convexité. On peut remarquer que l’intégrale (2.97) est formellement équivalente à une action d’Euler d’un fluide sans pression dont la dynamique est décrite par (2.85), mais sans le terme de gravité :

$$S = \int_{D=0}^1 dD D^{3/2} \int_{\mathbf{x}} d^3 \mathbf{x} \rho(\mathbf{x}) \|\mathbf{v}(\mathbf{x}, D)\|^2 \quad (2.98)$$

On peut discrétiser l’application  $\mathbf{q}(\mathbf{x})$  dans l’équation (2.97) en écrivant que  $\rho(\mathbf{x})$  est construit à partir de particules massives :

$$\rho(\mathbf{x}) = m_0 \sum_{i=0}^N \delta_D(\mathbf{x} - \mathbf{x}_i) \quad (2.99)$$

avec  $\mathbf{x}_i$  la position d’une des particules. L’ensemble  $\{\mathbf{x}_i\}$  constitue ce que l’on a appelé le maillage de MAK. L’application continue  $\mathbf{q}(\mathbf{x})$  est réduite à une application d’association  $\sigma$  entre les  $\mathbf{x}_i$  et les  $\mathbf{q}_{\sigma(i)}$ . Les  $\mathbf{q}_j$  sont distribuées de telle manière que, si les particules de masse étaient à ces positions, elles engendreraient un champ de densité homogène (à cette résolution). L’équation (2.97) peut donc être réécrite ainsi :

$$S_\sigma = m_0 \sum_{i=0}^N (\mathbf{x}_i - \mathbf{q}_{\sigma(i)})^2 \quad (2.100)$$

et la minimisation effectuée sur l’application  $\sigma$ . Une illustration du résultat d’un tel processus de minimisation sur un petit ensemble de particules est donnée figure 2.9. Il faut remarquer que cette méthode est quasiment identique à une autre s’appelant PIZA (Particle Interchange Zel’dovich Approximation, Croft & Gaztanaga, 1997). La différence subtile reside dans la formulation des propriétés recherchées du champ trouvé par la méthode, ainsi que dans l’algorithme utilisé pour l’obtenir (méthode stochastique pour PIZA, algorithme de “Auction” pour MAK, voir annexe B).



# 3 Test de la reconstruction MAK sur une simulation

## Sommaire

<b>3.1</b>	<b>La simulation <math>N</math> corps</b>	<b>29</b>
3.1.1	Les paramètres utilisés	29
3.1.2	Fluctuation à grande échelle du champ de vitesse	30
<b>3.2</b>	<b>Identification des halos</b>	<b>31</b>
<b>3.3</b>	<b>Reconstruction du champ de déplacement</b>	<b>31</b>
3.3.1	De la simulation	31
3.3.2	Sur un catalogue de halo	33
<b>3.4</b>	<b>Reconstruction du champ de vitesse</b>	<b>35</b>
3.4.1	Champ de vitesse de la simulation	35
3.4.2	Champ de vitesse sur FullMock	37
3.4.3	Statistique sur la direction	37
<b>3.5</b>	<b>Reconstruction des fluctuations primordiales</b>	<b>40</b>
3.5.1	Retrouver les conditions initiales de la simulation avec MAK	40
3.5.2	Le modèle du collapse sphérique	42
3.5.3	Application au champ de déplacement MAK	46
<b>3.6</b>	<b>Conclusion</b>	<b>47</b>

À PRÉSENT que nous savons en quoi consiste la méthode de reconstruction de Monge-Ampère-Kantorovitch (MAK) des champs de vitesse et que nous sommes en possession d'un algorithme de minimisation (annexe B), nous allons tester quelles sont les forces et les faiblesses de cette méthode à l'aide d'une simulation à  $N$  corps. Le but ici va être d'abord de retrouver le champ de vitesses de cette simulation. On s'intéressera d'abord (section 3.3) au champ de déplacement reconstruit. On utilisera pour cela des comparaisons entre des champ de vitesse lissés, soit les vitesses des "objets" eux-mêmes de la simulation. En deuxième temps (section 3.4), nous verrons comment obtenir une première analyse statistique des erreurs commises lors de la reconstruction du champ de vitesse. Enfin (section 3.5), nous nous intéresserons à la reconstruction des conditions initiales de cette simulation à l'aide de MAK.

## 3.1 La simulation $N$ corps

### 3.1.1 Les paramètres utilisés

Nous utilisons une simulation à  $128^3$  particules (Mohayaee et al., 2006) qui a été effectuée en utilisant la version publique du code HYDRA (Couchman et al., 1995) pour simuler la formation de structures par effondrement gravitationnel d'un ensemble de particules de matière noire décrites par des particules non-collisionelles. On suppose de plus que la formation des structures n'intervient que sous la forme de perturbations Newtoniennes par rapport à une métrique FLRW de type  $\Lambda$ CDM. La simulation couvre un volume comobile de  $200^3 h^{-3} \text{ Mpc}^3$ . La densité moyenne de matière est  $\Omega_m = 0.30$  et la constante cosmologique  $\Omega_\Lambda = 0.70$ . La constante de Hubble est  $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . La normalisation des fluctuations de densité dans une sphère de rayon  $8 h^{-1} \text{ Mpc}$  est  $\sigma_8 = 1.0$ . Nous faisons remarquer que cette valeur de  $\sigma_8$  est significativement plus grande que la valeur suggérée par les dernières données WMAP qui la place plutôt à  $\sigma_8 = 0.74$  (Spergel et al., 2006), mais ceci ne devrait pas trop affecter

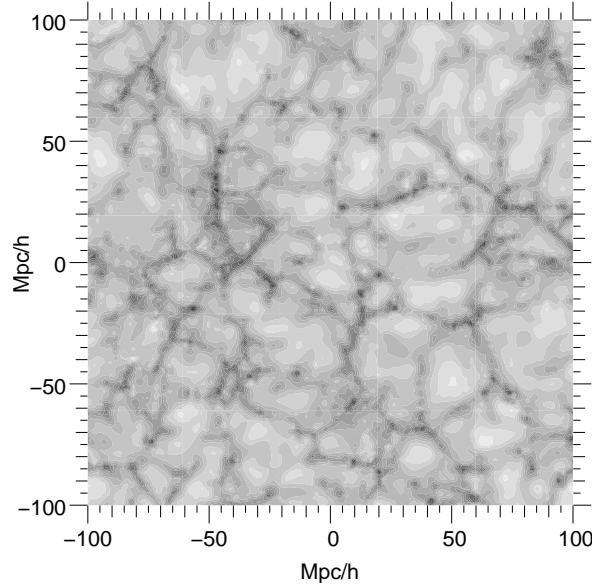


FIG. 3.1: Cette figure représente une coupe du champ de densité de la simulation à redshift  $z = 0$  (c'est-à-dire  $a = 1$ ). Le champ a été obtenu par filtrage adaptatif sur grille (appendice E) avec  $N = 32$  voisins. Les niveaux de gris sont proportionnel au logarithme du champ de densité. Les couleurs sombres donnent les régions de forte densité, et les couleurs claires les régions de faible densité.

les résultats présentés ici. En fait, un  $\sigma_8$  plus bas par rapport à 1.0 devrait réduire à la fois les non-linéarités et les effets de variance cosmique, ce qui devrait automatiquement améliorer la qualité de nos mesures.

Une coupe du champ de densité de cette simulation est présentée dans la figure 3.1.

### 3.1.2 Fluctuation à grande échelle du champ de vitesse

Comme le champ de vitesse fluctue de manière importante sur de grandes échelles, on pourrait s'inquiéter de la relative petite taille du volume de simulation. Nous allons présenter ici une estimation théorique de la fluctuation du champ de vitesse que l'on pourrait attendre sur une échelle de l'ordre de celle de la simulation. Nous allons pour cela nous appuyer sur la théorie linéaire énoncée dans la section 2.3. Supposons que nous connaissions le champ de vitesse  $\mathbf{u}(\mathbf{x}, t)$ . Nous allons à présent calculer une estimation de la variance en un point du champ de vitesse  $\mathbf{u}_W(\mathbf{x}, t)$  lissé à l'aide de la fonction  $W(\mathbf{x})$ . Le formalisme général des fonctions de corrélations lissées est donné appendice C. Nous allons ici l'appliquer au cas où  $W(\mathbf{x}) = W_{R, \text{tophat}}(\mathbf{x})$ . Il s'agit donc d'un filtrage de type top hat,<sup>1</sup> pour lequel on prendra un rayon de lissage tel que le volume de la sphère corresponde au volume de la simulation. On aura donc

$$\frac{4}{3}\pi R^3 = L^3 \implies R = \left(\frac{3}{4\pi}\right)^{1/3} L \simeq 0.62L. \quad (3.1)$$

En utilisant l'équation (C.18) avec cette fonction de filtrage et pour un rayon  $R = 124 h^{-1}\text{Mpc}$ , on trouve une dispersion de  $72 \text{ km s}^{-1}$ . Cette valeur doit être comparée aux erreurs caractéristiques qui vont apparaître lors de la reconstruction du champ de vitesse pour s'assurer que les effets de variance cosmique sont faibles en ce qui nous concerne.

<sup>1</sup>Filtrer un champ avec un top-hat consiste, pour obtenir le champ filtré en un point de l'espace  $\mathbf{x}$ , à moyenner la valeur du champ sur une boule de rayon précisé.

### 3.2 Identification des halos

Pour construire des catalogues virtuels, nous avons sélectionné des halos d’une simulation à  $N$  corps avec un algorithme bien connu appelé “Friend-Of-Friend” (FOF par la suite). Nous l’avons exécuté en utilisant une valeur traditionnelle du paramètre de liaison donné par  $l = 0.2$  (Efstathiou et al., 1988). Cette valeur permet de sélectionner des particules évoluant dans des régions dont la densité locale est supérieure à un certain seuil. Typiquement, si deux particules sont à une distance inférieure à  $L_{\text{liaison}} = l\bar{n}^{-1/3}$ , alors la densité locale de particule est de

$$n \simeq \frac{2}{L_{\text{liaison}}^3} = \frac{2}{l^3} \bar{n} \quad (3.2)$$

Donc, choisir  $l = 0.2$  consiste à isoler les régions où la densité en particules est 250 fois la densité moyenne de particules de la simulation (soit, de manière équivalente, 250 fois la densité moyenne de matière dans la simulation).

Les halos avec moins de 5 particules, c’est-à-dire avec une masse plus petite que  $M_{\text{min}} = 1.62 \times 10^{12} h^{-1} M_{\odot}$ , sont dissociés. La figure 3.2 montre un bon accord entre la fonction de masse des halos et le modèle de Sheth & Tormen (2002) pour les halos dont la masse est  $M \gtrsim M_{\text{min}}$ . Cependant, à peu près 63% de la masse n’est associée à aucun de ces halos et est distribuée dans un champ *d’arrière plan*. Dans des échantillons réalistes de galaxies comme le NBG-8k ou le catalogue 2MASS, la plus petite masse est de l’ordre de  $10^{11} M_{\odot}$ , valeur qui est sensiblement plus petite que notre  $M_{\text{min}}$ . Pour simuler des galaxies avec une masse plus petite que  $M_{\text{min}}$ , ce qui sera requis par la suite, nous utiliserons simplement les particules de matière noire qui n’ont pas été assignées à un halo comme des traceurs. Le catalogue contenant tous les halos et toutes les particules de *champ* sera appelé *FullMock*.

On pourrait s’inquiéter que la simulation à  $N$  corps que nous utilisons a une résolution trop basse pour étudier ce phénomène, car la distribution des petits halos est biaisée mais pas celle des particules de champ. Nous avons en fait vérifié qu’utiliser une simulation à  $512^3$  particules avec quasiment la même cosmologie [la simulation est décrite dans Colombi et al. (2007)] ne change pas les mesures présentées dans la section § 4.1.1.

### 3.3 Reconstruction du champ de déplacement

Deux types de reconstruction peuvent être effectuées. Le premier consiste à appliquer directement l’algorithme de MAK sur les particules d’une simulation à  $N$  corps. On obtient alors une reconstruction du déplacement de chacune de ces particules. Le deuxième est fondé sur l’utilisation du catalogue *FullMock* précédemment constitué pour reconstruire le déplacements des objets de ce catalogue.

#### 3.3.1 De la simulation

On va tester ici la reconstruction du champ de déplacement de la simulation en appliquant la procédure MAK aux particules de la simulation. Les deux éléments constituant la figure 3.3 permettent, dans une première approche qualitative, de comparer visuellement les résultats obtenus pour la première composante du champ de déplacement selon qu’on l’obtient par reconstruction ou par utilisation des vitesses des particules de la simulation. On constate à première vue que la reconstruction MAK est capable de capturer d’innombrables détails du champ de déplacement. Néanmoins on voit aussi que l’image du panneau de droite (champ de déplacement reconstruit) présent un aspect plus lissé que celle du panneau de gauche (champ de déplacement simulé). Par conséquent, il est probable que la reconstruction MAK n’est pas à même de capturer les détails à petites échelles, mais bien plus vraisemblablement la dynamique à grande échelle.

Une comparaison plus quantitative entre les deux champs est illustrée par les panneaux de la figure 3.4. Le panneau de gauche montre la comparaison entre le champ de déplacement simulé et celui reconstruit, sans lisser aucun des deux champs. L’accord est déjà fabuleux avec un coefficient de corrélation  $r = 0.9784$ . Le panneau de droite montre le résultat de la comparaison entre les deux



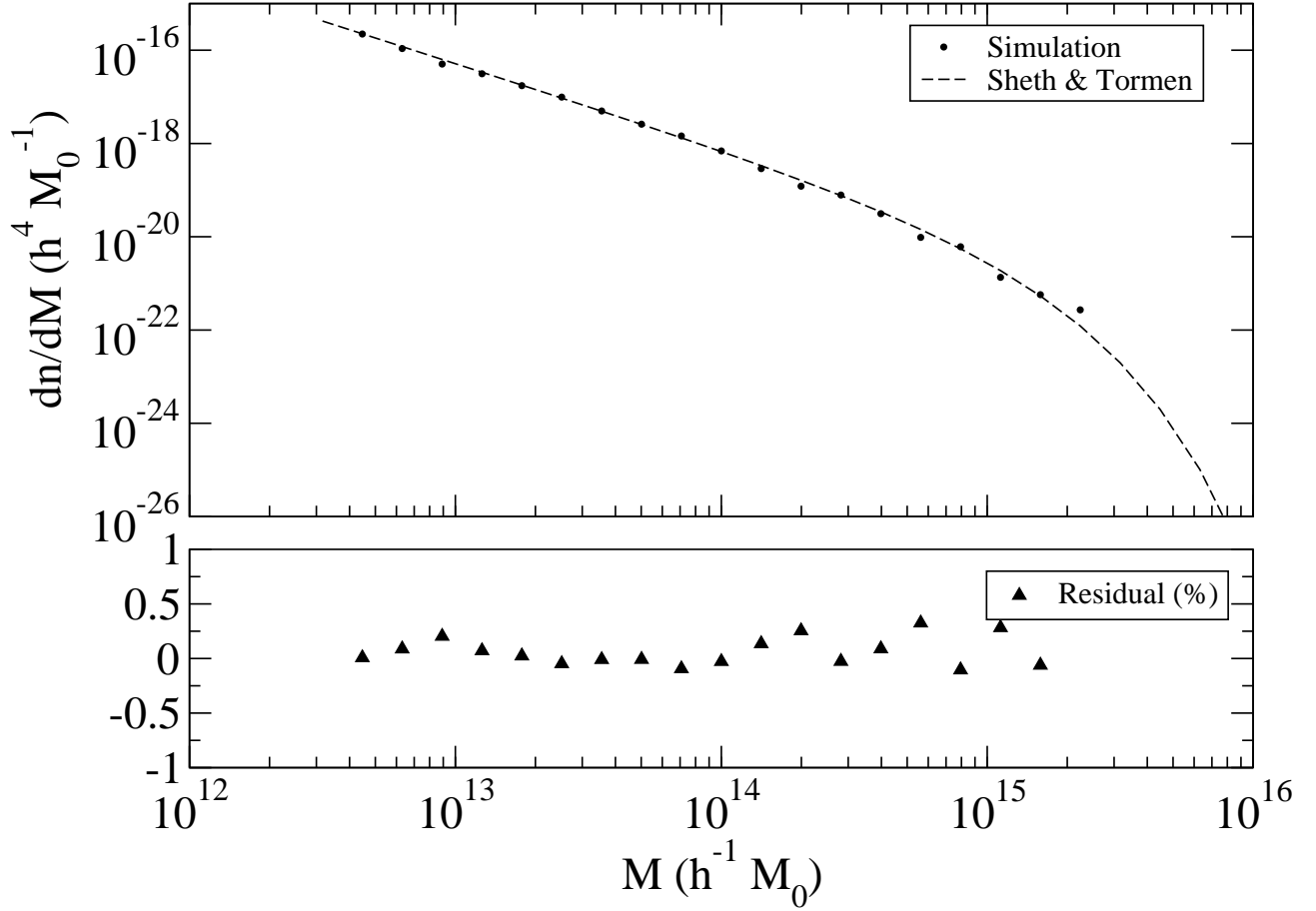


FIG. 3.2: *Fonction de masse de Sheth & Tormen / Masse diffuse* – Le panneau du haut donne la densité numérique d’halos dans un intervalle de masse en fonction de la masse. Les points ronds donnent la mesure de cette fonction dans FullMock alors que la ligne en tirets est obtenue en utilisant la théorie de Sheth & Tormen (2002) (cf. aussi Appendice D). Les résidus entre la prédiction et la mesure sont indiqués dans le panneau du bas (différences relatives). La plupart du temps, les points sont à quelques pour-cent de la prédiction théorique.

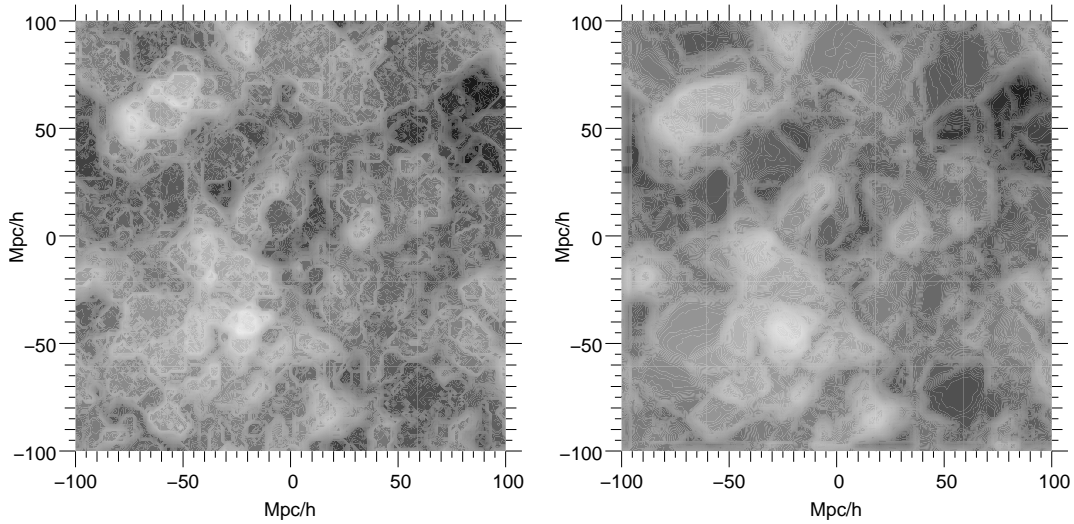


FIG. 3.3: Représentation de la première composante du champ de déplacement en coordonnées lagrangiennes pour la simulation (panneau de gauche) et la reconstruction (panneau de droite). (abscisse= $y$ , ordonnée= $z$ , blanc= $20 \text{ Mpc/h}$ , noir= $-20 \text{ Mpc/h}$ ). Les champs n'ont pas été lissés.

champs si on lisse avec un noyau gaussien de rayon  $5 h^{-1} \text{Mpc}$ . Cette fois on obtient quasiment une droite. Le coefficient de corrélation est de  $r = 0.996$ .

### 3.3.2 Sur un catalogue de halo

Cette fois nous allons essayer de faire une reconstruction sur un catalogue de halos (*FullMock*). Les halos détectés par l'algorithme Friend-Of-Friend sont considérés comme des objets ponctuels. Toutes les particules de chacun des halos sont placées au centre de masse de leur halo respectif. Le maillage de MAK correspondant a une résolution de  $128^3$  (identique à celle de la simulation). Dans cette partie, nous voudrions aussi tester l'effet de la résolution du maillage. Malheureusement, alors qu'il n'y a pas d'ambiguïté à initialiser un maillage  $128^3$  de MAK, quand on utilise tous les halos et les particules de champs (comme dans *FullMock*), il est moins trivial de considérer les résolutions inférieures qui seront utilisées dans quelques unes des analyses qui suivent, en particulier celles du chapitre 5. En effet, le nombre d'éléments du maillage assigné à chaque traceur n'est plus nécessairement un entier. Nous allons voir un algorithme plus général qui permet de créer des maillages de MAK de résolution arbitraire à partir d'un catalogue de halo. Pour tester les effets de la résolution sur la reconstruction du champ de vitesse, nous utiliserons cet algorithme pour produire un maillage de MAK à  $64^3$ .

#### Construction du maillage pour la reconstruction MAK

La reconstruction MAK requiert un échantillonnage de la distribution de matière avec des “particules” de même masse, qui correspondent aux noeuds d'un maillage uniforme en coordonnées lagrangiennes. Si on considère une simulation, on peut directement utiliser son maillage périodique complet. Cependant, dans des catalogues de galaxies (ou comme ici dans des catalogues de halos), le volume lagrangien est un sous-ensemble non périodique inscrit dans un sur-maillage parallélipédique. Dans ce cas, l'algorithme de reconstruction est appliqué uniquement à ces “particules” qui appartiennent à ce volume initial. Il faut remarquer que la détermination de ce volume initial est déjà par elle-même une tâche hautement non-triviale au cours de laquelle un mauvais choix peut avoir des conséquences dramatiques sur le champ de vitesse reconstruit.

Étant donné un ensemble de “galaxies”, ou des traceurs du champ de masse, pour lesquelles les masses individuelles  $M_i$  sont connues, et un choix de la résolution de masse pour la grille MAK,  $m_R$ ,

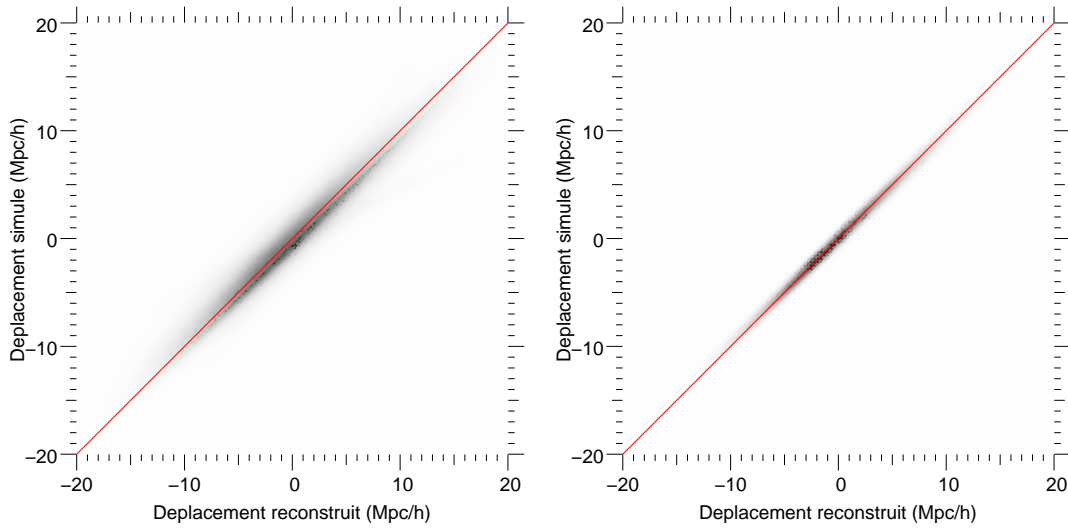


FIG. 3.4: Comparaison entre la première composante du champ de déplacement reconstruit (axe des ordonnées) et la première composante du champ de déplacement simulé (axe des abscisses). Le panneau de gauche représente la résultat de la comparaison sans avoir lissé les champs en coordonnées lagrangiennes. Le panneau de droite donne le résultat de la comparaison après lissage avec un noyau gaussien de  $5 h^{-1}\text{Mpc}$  en coordonnées lagrangiennes. La ligne rouge indiquerait le résultat d’une reconstruction parfaite.

le problème est maintenant de déterminer combien de “particules” doivent être assignées au traceur  $i$ . Ce nombre doit être  $n_i = M_i/m_R$ , qui n’est presque jamais un entier. Pour résoudre ce problème, nous allons construire une fonction entière  $\tilde{n}_i$  telle que la quantité

$$\chi^2 \equiv \sum_i \left( \frac{\tilde{n}_i m_R - M_i}{M_i} \right)^2 \quad (3.3)$$

soit minimale étant donnée la contrainte

$$\sum_i \tilde{n}_i = N_{\text{MAK}} , \quad (3.4)$$

avec  $N_{\text{MAK}}$  le nombre total de noeuds dans la grille MAK, telle que  $N_{\text{MAK}} \times m_R$  est aussi près que possible de la masse totale  $\sum M_i$ . La minimisation de cette quantité permet de minimiser l’erreur *relative* commise sur la masse et non l’erreur absolue. La minimisation du  $\chi^2$  est effectuée itérativement jusqu’à convergence à l’aide d’un algorithme de tri. Il faut remarquer que la solution donnée par une telle minimisation n’est pas, en général, unique à cause des permutations possibles entre les objets de même masse qui aboutit à un  $\chi^2$  équivalent.<sup>2</sup> A cause de cette dégénérescence, il est nécessaire de mélanger aléatoirement les traceurs avant d’effectuer la minimisation afin d’éviter des effets systématiques potentiels.

Il faut finalement remarquer que l’on doit s’assurer qu’il y a au moins quelques particules par traceur,  $n_i \geq \alpha$  avec  $\alpha > 1$ . Cela amène une contrainte sur  $m_R$  et par conséquent sur la taille (nombre de particules) du maillage de MAK. Malheureusement, il n’est pas toujours possible d’avoir  $\alpha > 1$  à cause du coût extrême en temps de calcul que cela impliquerait pour la reconstruction MAK pour les différents tests réalisés dans cette thèse. Pour résoudre ce problème, nous séparons le catalogue en groupes de galaxies et en galaxies de champ. Pour les groupes, la minimisation  $\chi^2$  est effectuée comme

<sup>2</sup>Cette remarque n’est vraie que dans le cas des simulations à  $N$ -corps où la masse est quantifiée par le nombre de particules. Dans les vraies observations, ce problème a peu de chance de se produire car la probabilité est nulle d’avoir deux objets dont la masse est totalement identique.

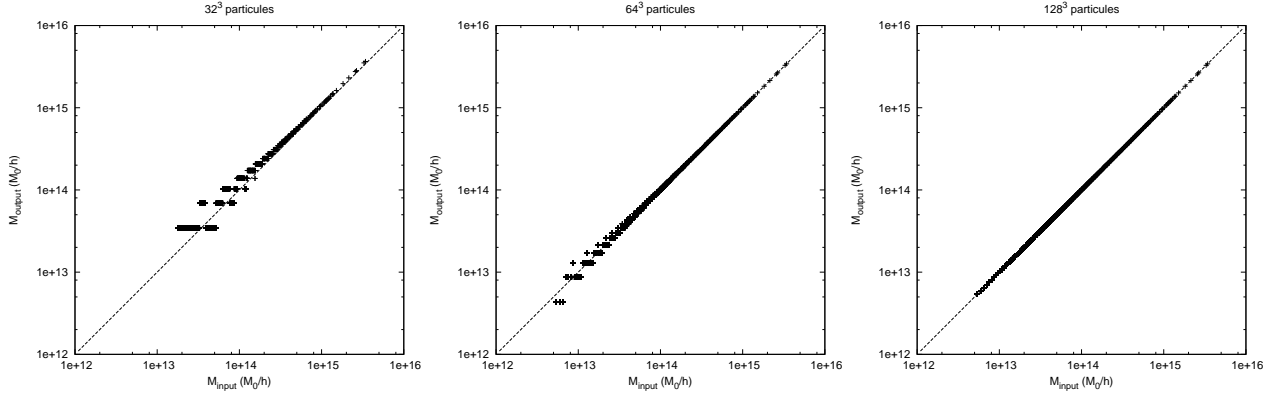


FIG. 3.5: *Construction d'un maillage de MAK* – Résultat donné par l'optimisation de la fonction  $\chi$  (équation (3.3)) pour la fonction donnant le nombre d'éléments du maillage MAK affecté à au halo  $i$   $\tilde{n}_i \times m_R$  versus la masse du halo  $i$ ,  $M_i$ . On a pris  $N_{\text{MAK}} = 32^3, 64^3$ , et  $128^3$  respectivement de gauche à droite.

il a été expliqué précédemment, avec la possibilité de perdre les groupes les plus légers car  $n_i$  peut être plus petit que un. Pour les galaxies de champ, nous utilisons une procédure plus simple : étant donné la masse  $M_i$  d'une galaxie  $i$ , un traceur MAK lui est affecté aléatoirement avec une probabilité d'occurrence  $M_i/m_R$ .

## Résultat

Les résultats de la reconstruction du champ de déplacement pour *FullMock* à une résolution de  $128^3$  sont donnés sur les figures 3.6 et 3.7. Le panneau de droite de la figure 3.6 montre l'aspect absolument remarquable de la reconstruction du champ de déplacement. Mise à part le fait qu'il semble plus lisse que le champ à gauche, il suit absolument toutes les caractéristiques du champ donné par la simulation, et ce sans lissage additionnel.

Un des avantages du catalogue FullMock est qu'il permet d'étudier la structure reconstruite du volume lagrangien des halos de la simulation. On peut observer sur la figure 3.7 que l'essentiel du volume projeté de ce halo massif ( $2.65 \cdot 10^{15} M_\odot$ ) est bien reconstruit, même si le détail des fluctuations du contour de ce volume n'est pas identique. La méthode MAK permettrait donc bien de mettre des contraintes sur l'origine de la masse des structures que l'on observe dans le ciel.

## 3.4 Reconstruction du champ de vitesse

Une fois le champ de déplacement  $\Psi(\mathbf{q})$  trouvé, on peut utiliser la relation (2.75) (page 18) issu de l'approximation de Zel'dovich pour trouver la vitesse d'un objet ou bien le champ de vitesse. En effet, nous rappelons que nous avons alors :

$$\mathbf{v}(\mathbf{x}) = \mathbf{v}(\mathbf{q}(\mathbf{x})) = H_0 \beta \Psi(\mathbf{q}(\mathbf{x})), \quad (3.5)$$

avec  $H_0$  la constante de Hubble à l'instant présent et  $\beta \simeq \Omega_m^{5/9}$  pour une cosmologie  $\Lambda$ CDM. Nous allons à présent étudier quelles sont les succès et les limites de l'utilisation d'une telle relation en synergie avec la méthode de reconstruction d'orbite MAK pour obtenir les champs de vitesse.

### 3.4.1 Champ de vitesse de la simulation

Comme précédemment sur la simulation, nous nous penchons à présent sur les résultats de la reconstruction du champ de vitesse lui-même. Sur la figure 3.8, on a représenté la composante suivant la

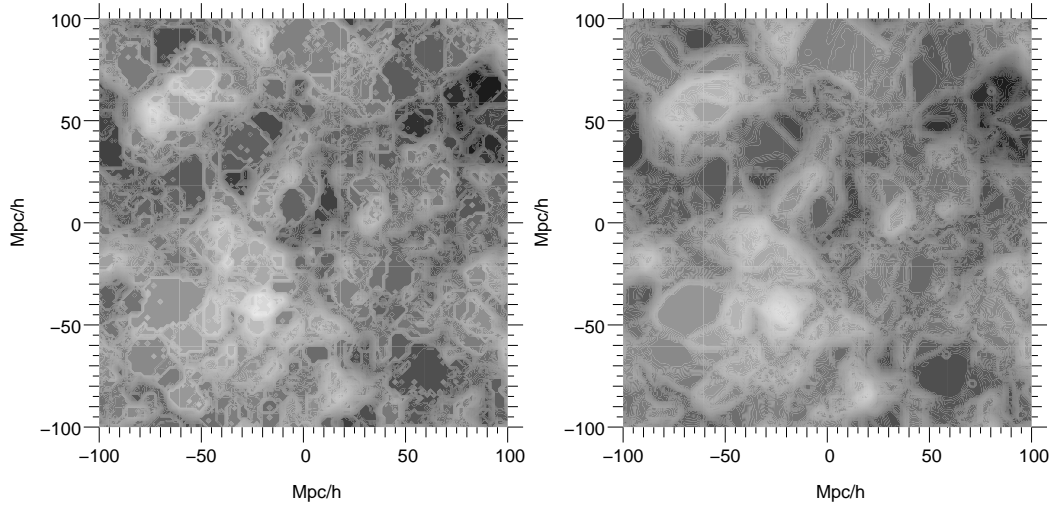


FIG. 3.6: *Champ de déplacement de FullMock* – Coupe de la première composante du champ de déplacement exprimé en coordonnées lagrangiennes, tel que donné par la simulation (à gauche) et par la reconstruction (à droite). Les deux champs n’ont pas été lissés.

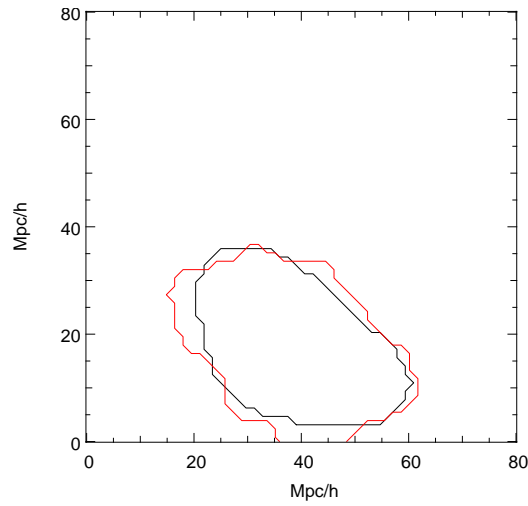


FIG. 3.7: Volume lagrangien projeté selon l’axe Z d’un groupe du catalogue FullMock. Le contour rouge donne le volume tel que donné par la simulation. Le contour noir est reconstruit par MAK.

ligne de visée du champ de vitesse de la simulation lissé avec un noyau gaussien de largeur  $5 h^{-1}\text{Mpc}$  ainsi que le champ reconstruit, en supposant que l’observateur est au centre de la boîte. Nous commençons à introduire le concept d’observateur car toutes les mesures ne pourront être obtenues que par comparaison à la seule composante “radiale” (ou suivant la ligne de visée). L’accord entre ces deux champs est tout à fait remarquable sur l’ensemble du champ présenté. On peut s’en rendre compte tout particulièrement sur le panneau du bas de la figure 3.8 qui met en évidence une très bonne relation linéaire entre le champ de vitesse reconstruit et celui simulé (l’échelle des niveaux de gris est logarithmique). Il existe néanmoins des régions où ce n’est pas le cas comme c’est mis en avant par les queues de cette même distribution. En effet, ces queues de distribution non-triviales se situent au-dessus de la diagonale dans le quadrant en bas à gauche et en dessous pour le quadrant en haut à droite. On dit que la forme d’une telle distribution de probabilité (PDF) est une hélice. Si on observe attentivement les deux champs représentés juste au-dessus, on ne constate en effet que les vitesses élevées en valeur absolue tendent à être sous-estimées. Ceci est particulièrement vrai par exemple pour le pic blanc brillant en bas à gauche du panneau milieu-gauche de la figure 3.8. Si on compare à quoi sont liées ces différences sur le champ de densité on constate une probable corrélation avec les pics du champ de densité. Cela signifie que les régions très surdenses ont tendance à avoir un champ de vitesse mal reconstruit. On pouvait s’y attendre car, dans ce cas, on tend à étudier une dynamique dont les non-linéarités n’ont pas été capturées par notre méthode mélangeant la reconstruction MAK et l’approximation de Zel’dovich. Ces queues de distributions sont susceptibles de produire un petit biais lors des comparaisons vitesse-vitesse.

### 3.4.2 Champ de vitesse sur FullMock

Les résultats, donnés sur la figure 3.9, pour la reconstruction  $128^3$  sont quasiment inchangés par rapport à ceux sur la simulation. Néanmoins, nous avons à présent accès à une comparaison des vitesses objets pris individuellement (panneau gauche-bas). On constate que cette distribution a le même problème que celle du champ de vitesse. Elle prend en effet une forme d’hélice aux deux extrémités.

Nous testons ici pour la première fois l’effet d’un choix de résolution inférieure pour la reconstruction à celle de la simulation. Si on compare les panneaux du milieu à ceux de gauche, on constate quasiment aucune différence, notamment pour le champ de vitesse. La largeur du diagramme de dispersion de la reconstruction  $64^3$  (panneau du milieu-bas) est légèrement supérieure à celle de la  $128^3$ . Cela peut s’expliquer par le fait qu’on utilise une grille plus grossière pour la distribution des particules dans les coordonnées lagrangiennes. Nous y reviendrons avec la figure 4.1 à la section 4.1.1 (page 51).

### 3.4.3 Statistique sur la direction

Une autre quantité statistique importante est de savoir, pour un objet ou un point donné de l’espace, quel est l’ordre de grandeur de l’erreur sur la direction du champ de vitesse reconstruit. Pour cela, nous avons mesuré la quantité suivante

$$\alpha = \cos^{-1} \left( \frac{\mathbf{v}_{\text{rec}} \cdot \mathbf{v}_{\text{sim}}}{\|\mathbf{v}_{\text{rec}}\| \|\mathbf{v}_{\text{sim}}\|} \right), \quad (3.6)$$

avec  $\mathbf{v}_{\text{rec}}$  le vecteur vitesse reconstruit,  $\mathbf{v}_{\text{sim}}$  le vecteur vitesse simulé,  $\alpha$  l’angle entre ces deux vecteurs. Nous avons ensuite fait un histogramme donnant la densité de probabilité  $P(\alpha)$  d’avoir un vecteur vitesse reconstruit, soit un objet de FullMock soit le champ de vitesse reconstruit lissé adaptativement, avec une mauvaise direction  $\alpha$  par rapport au vecteur vitesse simulé. Nous avons représenté sur la figure 3.10 la probabilité cumulée  $P(< \alpha)$  que la vitesse soit reconstruite avec une erreur sur la direction inférieur à  $\alpha$  pour trois cas :

- si on considère l’ensemble de tous les objets (halos et particules non liées) de FullMock (courbe noire du panneau de gauche), on trouve alors que l’on a une probabilité de 95% d’avoir une erreur inférieure à 77 degrés ;

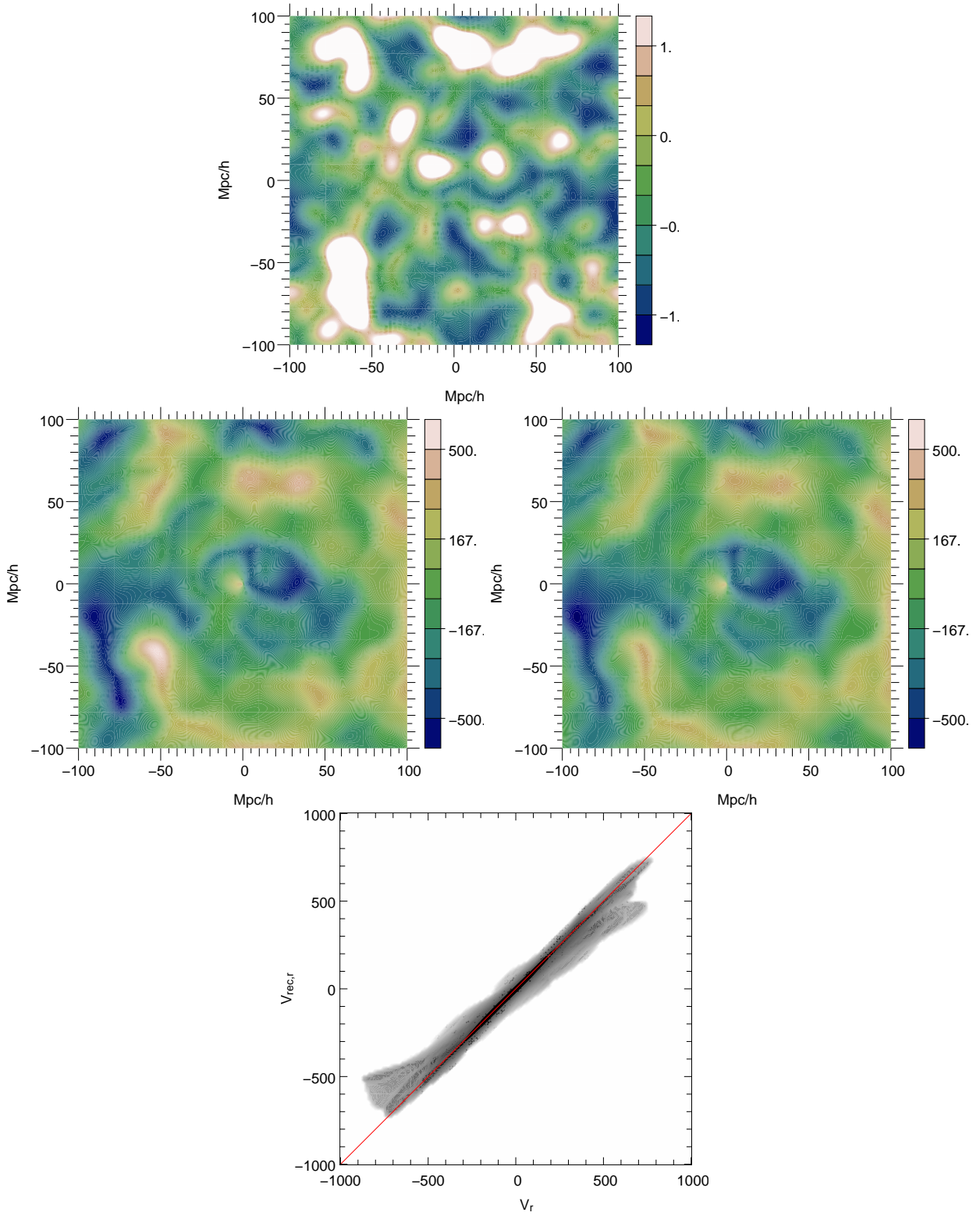


FIG. 3.8: *Reconstruction du champ de vitesse de la simulation* – Panneau du haut : Coupe du champ de densité de la simulation lissé avec un noyau gaussien de rayon  $5 h^{-1}\text{Mpc}$ . Panneau au milieu à gauche : Composante selon la ligne de visée du champ de vitesse de la simulation lissé avec un noyau gaussien de rayon  $5 h^{-1}\text{Mpc}$ . On a placé l’observateur en  $(0,0)$ . Panneau au milieu à droite : Même chose que pour le panneau de gauche mais on considère ici le champ de vitesse reconstruit. Panneau du bas : comparaison entre le champ de vitesse reconstruit (ordonnée) et simulé (abscisse). On a utilisé une échelle logarithmique de niveau de gris pour mettre en évidence les queues de distribution.

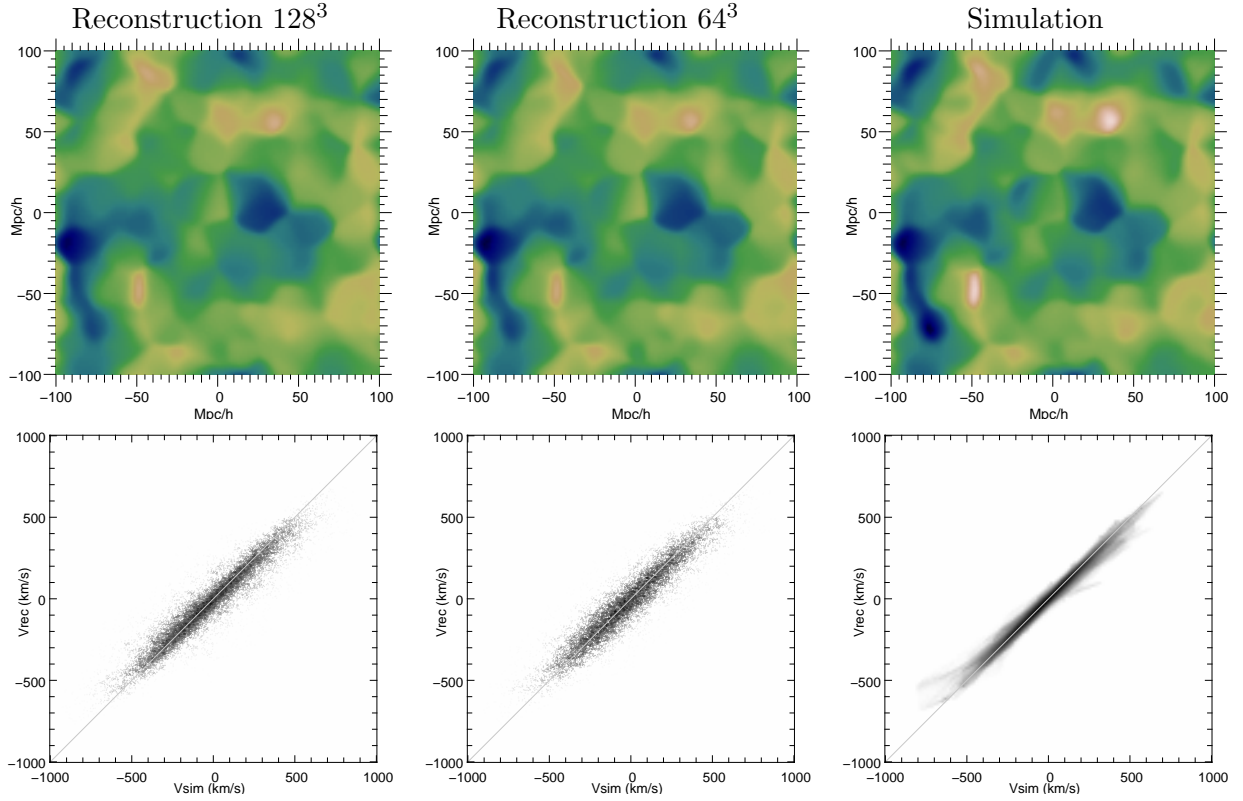


FIG. 3.9: *Reconstructin du champ de vitesse sur FullMock* – Panneaux du haut : Une coupe de la composante sur la ligne de visée du champ de vitesse simulé,  $v_{r,\text{sim}}$ , et celui reconstruit,  $v_{r,\text{rec}}$ , après lissage avec un fenêtre gaussienne de  $5 h^{-1}\text{Mpc}$  de rayon. L’observateur est au centre de cette coupe. Panneaux du bas : Diagramme de dispersion entre  $v_{r,\text{sim}}$  et  $v_{r,\text{rec}}$  pour des halos individuels (gauche) et après lissage (droite).



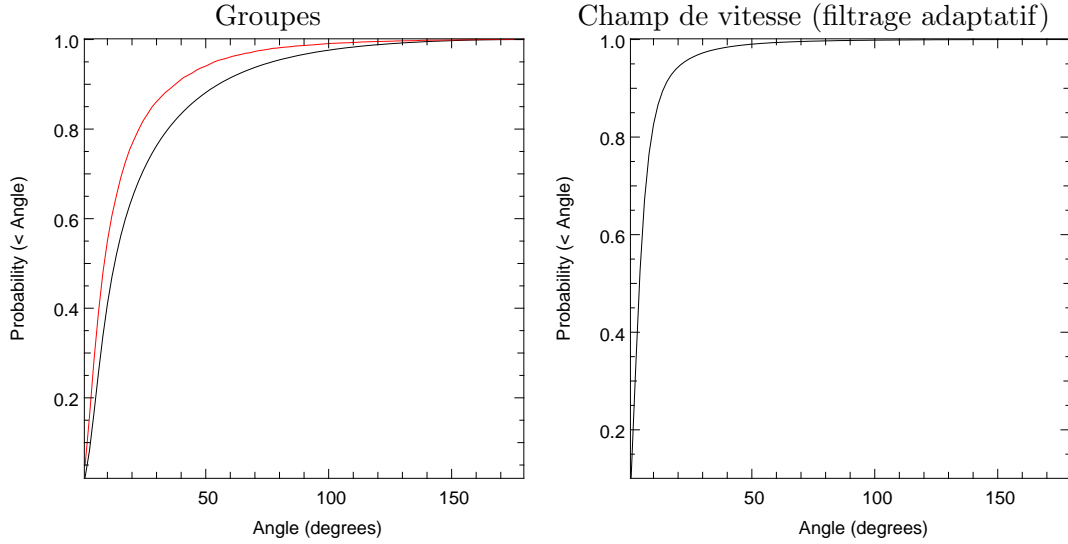


FIG. 3.10: *Distribution cumulée d'erreur de reconstruction de la direction du champ de vitesse* – On a représentée ici la distribution cumulée normalisée des erreurs de reconstruction sur la direction du champ de vitesse, obtenu par filtrage adaptatif dans le panneau de droite (annexe E) en un point donné du volume de la simulation et sur la vitesse des groupes de particules eux-même à gauche. Sur le panneau de gauche on distingue deux courbes : en noir, il s'agit de la distribution cumulée sur *tous les objets* (y compris ceux à une particule), en rouge, il s'agit de la distribution cumulée sur *seulement les objets composites* (particules regroupées par l'algorithme friend-of-friend). Il y a une probabilité de 95% d'avoir une erreur sur la direction de reconstruction inférieur à 22 degrés sur le champ de vitesse en un point de l'espace. Cet angle vaut 77 degrés pour la courbe noire du panneau de gauche et 53 degrés pour la courbe rouge.

- si on considère uniquement les halos de FullMock (courbe rouge du panneau de gauche), on trouve alors que l'on a une probabilité de 95% d'avoir une erreur inférieure à 55 degrés ;
- si on considère le champs de vitesse lissé adaptativement, alors la distribution est beaucoup plus étroite, on trouve que une probabilité d'erreur de 95% correspond à 22 degrés seulement.

Les erreurs que nous avons identifiées ici nous servirons à mettre une barre d'erreur sur la direction du champs de vitesse par la suite (notamment à la section 6.3).

## 3.5 Reconstruction des fluctuations primordiales

### 3.5.1 Retrouver les conditions initiales de la simulation avec MAK

La reconstruction MAK est capable de reconstituer les orbites des objets (galaxies ou groupes de galaxies) d'un catalogue à partir de leur position actuelle. En effet, on peut utiliser l'équation (2.67) (page 17) donnant perturbativement l'expression du champ de fluctuation de densité à tout instant dans le passé si on connaît les trajectoires  $\Psi(\mathbf{x}, t)$ . Dans le cadre de l'approximation de Zel'dovich ces fluctuations s'expriment sous la forme suivante :

$$\delta_{\text{NL}}(\mathbf{q}, t) = -\frac{D(t)}{D(t_0)} \text{div}_{\mathbf{q}} \Psi_0(\mathbf{q}) \quad (3.7)$$

avec  $\Psi_0(\mathbf{q})$  le champ de déplacement en coordonnées lagrangiennes à partir des coordonnées lagrangiennes jusqu'aux positions actuelles,  $D(t)$  le taux de croissance linéaire,  $t_0$  l'instant présent. L'indice "NL" sera discuté par la suite. Dans la pratique, les conditions initiales sont toujours normalisées en

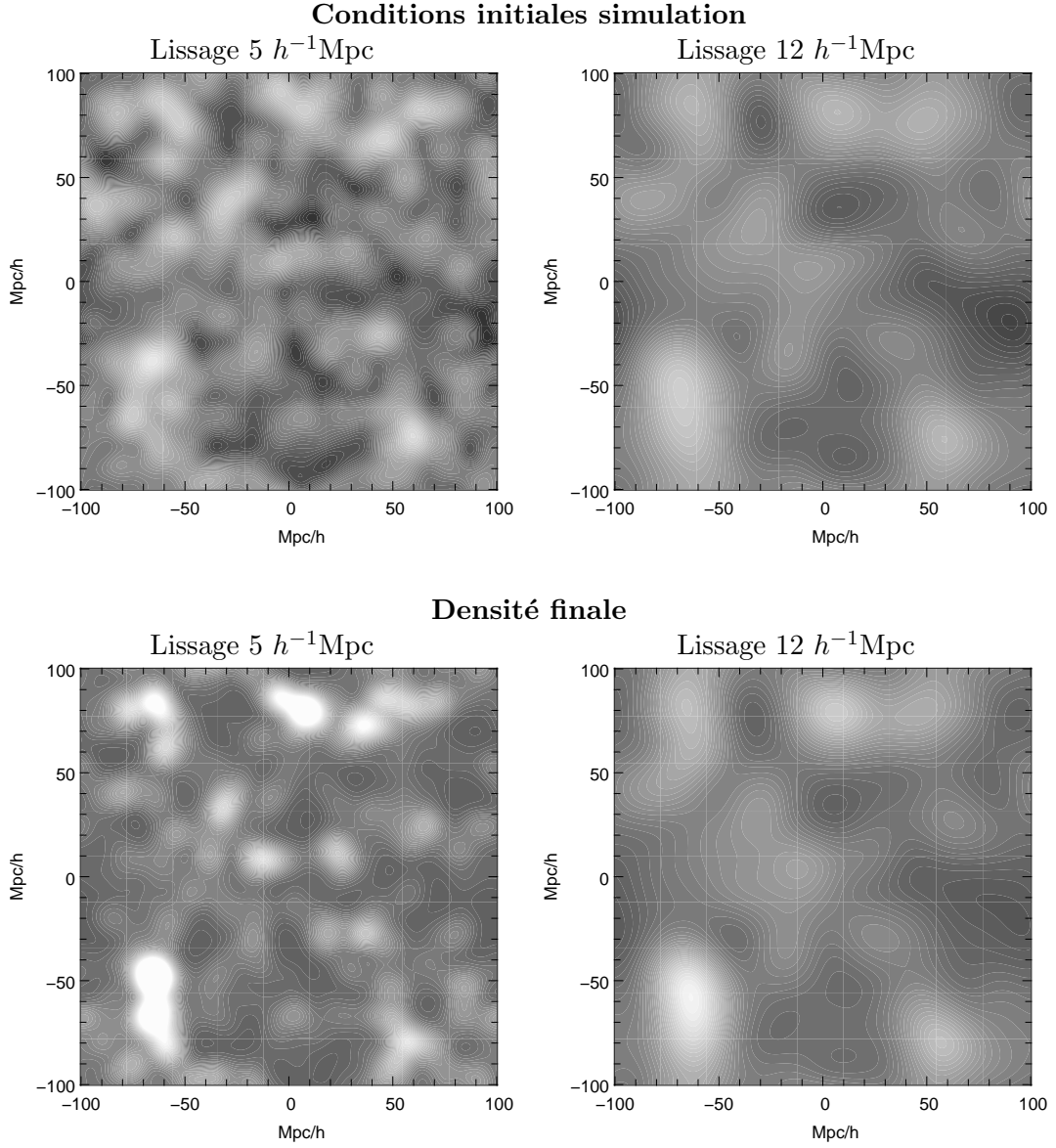


FIG. 3.11: *Tentative de reconstruction des conditions initiales avec la théorie linéaire* – Nous illustrons ici une tentative de reconstruction des conditions initiales à l’aide de la théorie linéaire seule. Les panneaux du haut donnent les conditions initiales lissés avec un noyau gaussien de rayon  $5 h^{-1}\text{Mpc}$  (gauche) et  $12 h^{-1}\text{Mpc}$  (à droite). En bas, nous avons effectué la même opération mais avec le champ de densité finale. Nous voyons qu’il faut au moins lisser avec à  $12 h^{-1}\text{Mpc}$  pour obtenir un champ “prédit” ressemblant aux conditions initiales de la simulation.

les extrapolant par la théorie linéaire à l’instant présent. Nous allons donc poser

$$\delta_{\text{NL}}(\mathbf{q}) = -\text{div}_{\mathbf{q}} \Psi_0(\mathbf{q}) \quad (3.8)$$

le champ reconstruit de conditions initiales de la simulation. Nous utiliserons le champ de déplacement reconstruit par MAK dans cette formule ci-dessus pour en inférer  $\delta_{\text{NL}}$ . C’est l’approche suggérée par Mohayaee et al. (2006). Nous présentons les résultats donnés par cette méthode sur les figure 3.12 et 3.13. Visuellement, le champ de conditions initiales reconstruit sans lissage (panneau en haut à droite de la figure 3.12) présente la même topologie que le champ de conditions initiales de la simulation. Néanmoins, il semble beaucoup plus filamenteux. Si on compare avec un champ de conditions initiales obtenu en utilisant le champ de déplacement de la simulation (panneau en haut au milieu), on constate cet aspect filamentaire est aussi présent et qu’en fait ce champ est quasiment le même que celui correspondant à la reconstruction. Comme on l’a déjà remarqué dans les sections précédentes, la méthode MAK reconstruit les orbites complètes des galaxies comme des trajectoires rectilignes. Malheureusement pour nous dans ce cas, cette orbite a intégré tous les effets non-linéaires qu’a ressentie la matière noire pour arriver à l’état courant (d’où l’indice “NL” dans les équations ci-dessus). Pour avoir accès à l’orbite dans le régime *linéaire* il nous faudrait donc les retrancher ou considérer des échelles plus grandes. On peut pour cela regarder la série de panneaux du bas de la figure 3.12, qui correspondent au même champs des panneaux du dessus mais cette fois lissés avec un noyau gaussien à  $5 h^{-1}\text{Mpc}$ . Cette fois, on voit bien que les trois champs se ressemblent, notamment ceux qui nous intéressent à savoir le champ issu du champ de déplacement reconstruit et le champ de conditions initiales de la simulation. On a représenté figure 3.13, panneau de gauche, la distribution de probabilité jointe d’avoir le couple  $(\delta_{\text{NL}}, \delta_{\text{simu}})$  en un point de l’espace lagrangien. On voit en effet une très bonne corrélation. Néanmoins, on observe aussi ce qui semble être un effet systématique, qui est plus prononcé dans le quadrant en bas à gauche. Cet effet semble moins prononcé dans le quadrant en haut à droite mais cela peut-être du à une relative plus grande largeur de la distribution. Une méthode de comprendre les non-linéarités supplémentaires introduites dans le champ de déplacement par rapport à l’approximation de Zel’dovich est d’étudier la dynamique de l’effondrement sphérique.

### 3.5.2 Le modèle du collapse sphérique

Le modèle du collapse sphérique est un modèle jouet mais qui permet déjà d’obtenir des résultats intéressant sur la dynamique des grandes structures dans un Univers en expansion. Il consiste à étudier l’évolution d’une surdensité sphérique dans un Univers  $\Lambda\text{CDM}$ . Nous supposons de plus que nous n’étudions que la dynamique des couches “externes”, c’est-à-dire celles pour lesquelles il n’y pas de multivaluation du champ de vitesse.

Nous partons de l’équation d’évolution d’une sphère homogène de particules de matière noire. En coordonnées physiques, l’équation d’une couche à la distance  $R$  est donnée par

$$\ddot{R} = -\frac{GM(< R)}{R^2} + \frac{8\pi G}{3}R\rho_{\Lambda} \quad (3.9)$$

avec  $\rho_{\Lambda}$  la densité d’énergie sombre,  $G$  la constante de gravité universelle et  $M(< R)$  la masse à l’intérieur de la sphère de rayon  $R$ . Comme nous étudions une couche “externe” nous pouvons supposer qu’à tout instant  $M(< R) = M_0$ . Nous allons paramétriser le temps à l’aide du facteur d’échelle  $a$  et nous placer en coordonnées comobiles :  $R(t) = R_0 a(t) b(a(t))$ , avec  $R_0$  le rayon physique de la coquille. La fonction  $b(a(t))$  va décrire l’évolution de l’effondrement par rapport à l’expansion de l’Univers, elle est censée ne pas comporter de singularités.

Avec une cosmologie  $\Lambda\text{CDM}$ , nous rappelons que (équation de Friedmann (2.12))

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\bar{\rho} - 2\rho_{\Lambda}). \quad (3.10)$$

On peut alors réécrire l’équation (3.9) sous la forme suivante :

$$2a^2 b''(a) = -\Omega_{\text{m}}(a) \frac{1 + \delta_0}{b^2} + ab'(a) (\Omega_{\text{m}}(a) - 2\Omega_{\Lambda}(a) - 4) + \Omega_{\text{m}}(a)b \quad (3.11)$$

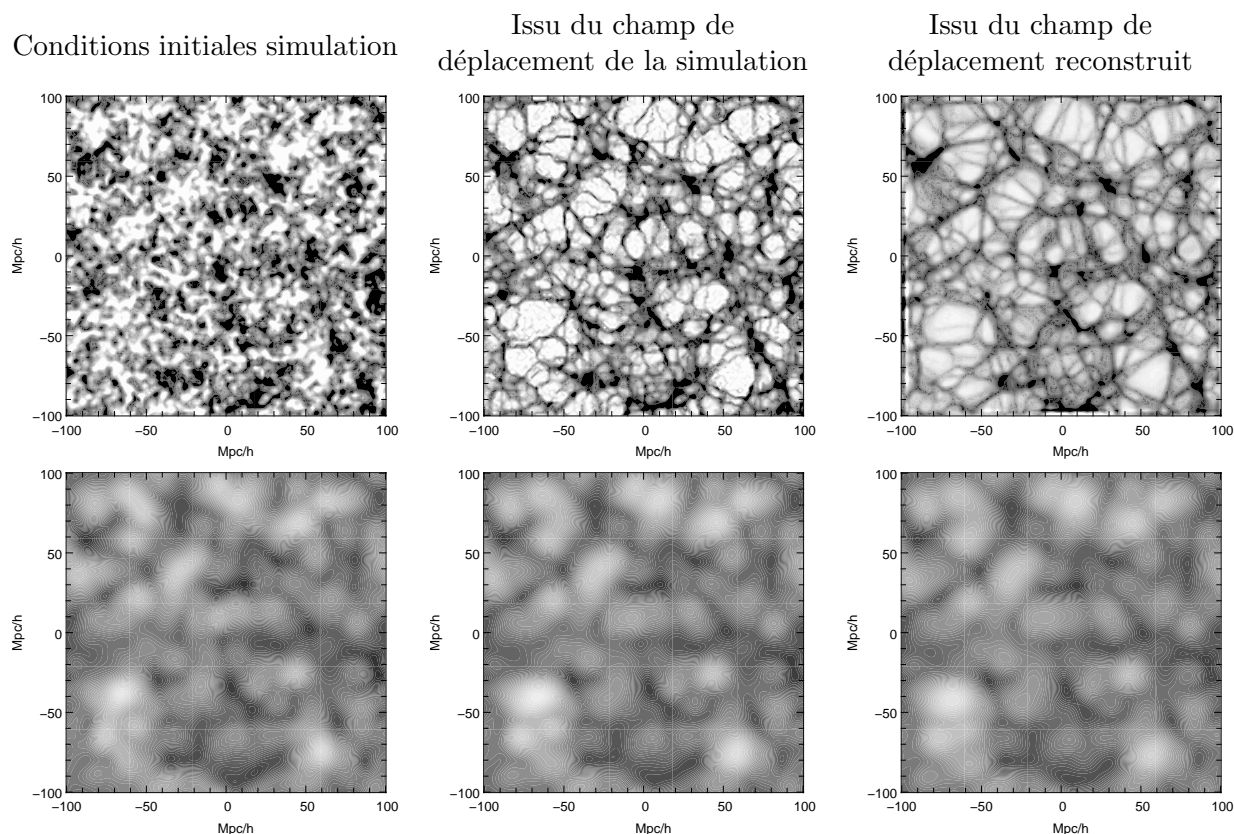


FIG. 3.12: *Reconstruction des fluctuations primordiales* – Illustration des champs des conditions initiales obtenues par différentes méthodes. Les champs de fluctuations ont tous été extrapolés linéairement à redshift nul. L'échelle des niveaux de gris a été choisie identique pour tous les panneaux en allant d'une fluctuation de  $-3$  (noir) à  $+3$  (blanc). **Première ligne de panneaux :** champ de densité non-lissé obtenu à l'aide de la relation (3.8) pour différents champs de déplacement : déplacement initial des particules avant d'exécuter la simulation à  $N$ -corps (panneau de gauche), déplacement total des particules obtenus après exécution de la simulation (panneau du milieu), déplacement reconstruit par la méthode de reconstruction MAK (panneau de droite). **Seconde ligne de panneaux :** Les mêmes champs de densité que ci-dessus, mais lissés avec un noyau gaussien de rayon  $5 h^{-1}\text{Mpc}$ .

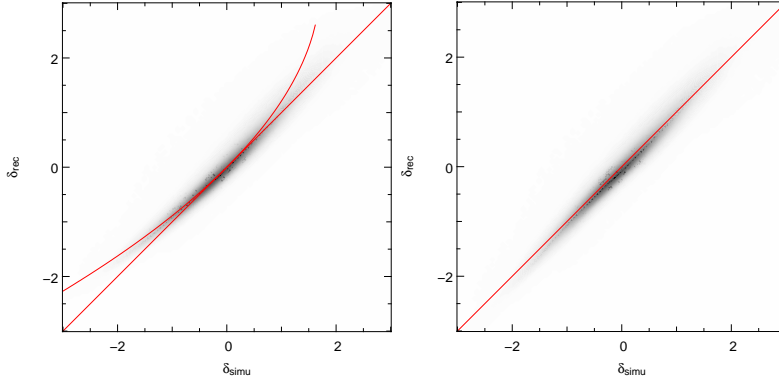


FIG. 3.13: Comparaison du champ de conditions initiales  $\delta_{\text{rec}}$  reconstruites à celles données à la simulation, une fois que ces deux champs ont été lissés avec un noyau gaussien de rayon  $5 h^{-1}\text{Mpc}$ . Le graphe de gauche montre le résultat de la comparaison sans correction de l'effondrement sphérique. On a ajouté à ce diagramme une courbe en rouge correspondant au résultat donné par la théorie de l'effondrement s. Le graphe de droite montre le résultat après correction de l'effondrement sphérique pour  $\delta < 0$  uniquement.

avec

$$\begin{aligned}\Omega_m(a) &= \frac{\bar{\rho}}{\rho_c} = \frac{\Omega_{m,0}}{\Omega_{m,0} + a^3 \Omega_\Lambda}, \\ \Omega_\Lambda(a) &= \frac{\rho_\Lambda}{\rho_c} = 1 - \Omega_m(a),\end{aligned}\quad (3.12)$$

et  $\delta_0$  la surdensité initiale dans la sphère de rayon comobile  $R_0$ . On peut tout de suite voir que c'est une équation qui n'est, en général, pas aisément soluble.

Nous allons étudier des faisceaux de solutions pour  $\Omega_m \leq 1$ . Pour cela nous allons spécifier les conditions initiales de manière à ce que la croissance du contraste de densité suive le mode croissant de la théorie linéaire Eulérienne. Pour cela la vitesse comobile  $\mathbf{u}$  de la couche est

$$\mathbf{u} = \left( \frac{dR}{dt} - HR(t) \right) \hat{\mathbf{x}} = R_0 a \dot{a} b'(a) \hat{\mathbf{x}} \quad (3.13)$$

D'après l'équation (2.63), on a

$$\text{div}_{\mathbf{x}} \mathbf{u} = -\dot{a} \beta \delta(\mathbf{x}, t), \quad (3.14)$$

ce qui implique que la relation dans les conditions initiales est :

$$-\dot{a} \beta \delta_0 = \text{div}_{\mathbf{x}} (R_0 a_0 \dot{a}_0 b'(a_0) \hat{\mathbf{x}}) = 3a_0 \dot{a}_0 \frac{b'(a_0)}{b(a_0)}, \quad (3.15)$$

pour  $a_0$ ,  $\dot{a}_0$  le facteur d'expansion et la dérivée temporelle du facteur d'expansion dans les conditions initiales du collapse sphérique. Mais si on impose que  $b(a_0) = 1$  (le rayon  $R_0$  est le rayon initial de la fluctuation) alors on a

$$R_0 a_0 \dot{a}_0 b'(a_0) = -a_0 H_0 \delta_0 \frac{R_0}{3}. \quad (3.16)$$

En définitive, les conditions initiales que l'on doit choisir sont :

$$\begin{aligned}b(a_0) &= 1 \\ b'(a_0) &= -\frac{\delta_0}{3a_0}\end{aligned}\quad (3.17)$$

On a représenté sur le graphique de la figure 3.14, un exemple de solution obtenue numériquement pour  $b(a)$ , avec  $\delta_0 = 10^{-3}$ ,  $a_0 = 10^{-3}$ .

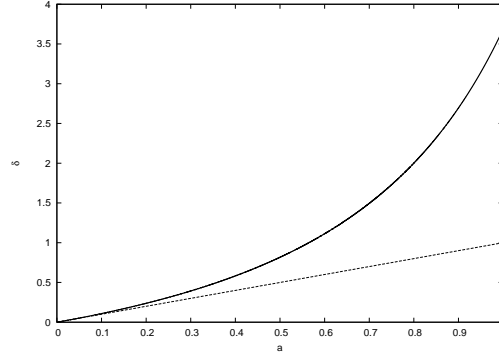


FIG. 3.14: Exemple d'évolution du paramètre de surdensité  $\delta = (\rho - \bar{\rho})/\bar{\rho}$  d'une surdensité sphérique dans un Univers Einstein-de-Sitter. La ligne pleine donne l'évolution donnée par la résolution numérique de l'équation (3.11) pour  $\delta_0 = 10^{-3}$ ,  $a_0 = 10^{-3}$ ,  $\Omega_{m,0} = 1$ ,  $\Omega_{\Lambda,0} = 0$ . La ligne en tiret donne l'évolution de la surdensité telle que prédite par la théorie linéaire eulérienne.

### Retour au champ de fluctuations initiales : utilisation du collapse sphérique

A présent que nous avons un modèle d'évolution simplifié d'une surdensité sphérique nous pouvons nous en servir pour corriger statistiquement la relation entre

$$\delta_{\text{NL}} \equiv -\text{div}_{\mathbf{q}} \Psi(\mathbf{q}) \quad (3.18)$$

et  $\delta_{\text{L}}$  le champ des fluctuations primordiales *linéaire* normalisé à  $z = 0$ . Comme on l'a précisée dans la section 3.5.1, l'approximation de Zel'dovich (2.64) devrait donner

$$\delta_{\text{NL}} \simeq \delta_{\text{L}} . \quad (3.19)$$

Mais il probablement possible d'avoir une meilleure reconstruction des fluctuations primordiales si on tient compte du fait qu'un petit volume d'Univers  $\Delta^3 \mathbf{q}$  placé à la coordonnée lagrangienne  $\mathbf{q}$  peut légèrement s'effondrer sur lui-même et que cette dynamique n'est pas bien prise en compte par la reconstruction MAK.

Nous cherchons un lien entre  $\delta_{\text{L}}$  et  $\delta_{\text{NL}}$  dans le cadre de l'effondrement sphérique. Pour cela nous partons de la définition de  $\delta_{\text{NL}}$ .

$$\delta_{\text{NL}} \equiv -\text{div}_{\mathbf{q}} \Psi . \quad (3.20)$$

Comme le mouvement des particules de matière noires, dans le cadre de ce modèle, est purement radial, on peut écrire en coordonnées comobiles :

$$\Psi(\mathbf{q}, a) = R_0(b(a) - 1)\hat{q} = (b(a) - 1)\mathbf{q}, \quad (3.21)$$

d'où l'expression de  $\delta_{\text{NL}}$  :

$$\delta_{\text{NL}} = -3 \left( (1 + \delta)^{-1/3} - 1 \right), \quad (3.22)$$

car  $b(a) \simeq 1/(1 + \delta(a))^{-1/3}$  par conservation de la masse, avec  $\delta$  la surdensité non-linéaire due à la dynamique de l'effondrement. Pour mémoire,  $\delta$  a été représentée sur la figure 3.14. Bernardeau (1994) argumente que la valeur finale de  $\delta$  (notée ici  $\delta_{\text{f}}$ ) pour  $a = 1$  est reliée d'une manière très simple à  $\delta_{\text{L}}$ , la valeur de la surdensité initiale extrapolée à l'aide de la théorie linéaire à l'instant présent  $a = 1$  :

$$1 + \delta_{\text{f}} \simeq \left( 1 - \frac{2}{3}\delta_{\text{L}} \right)^{-3/2}. \quad (3.23)$$

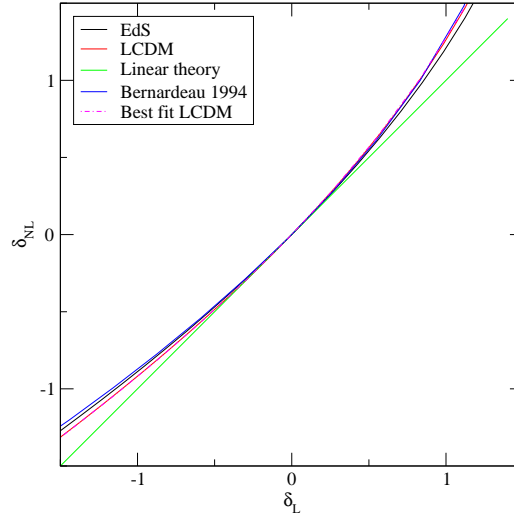


FIG. 3.15: *Collapse sphérique et conditions initiales* – On a représenté ici la prédiction de la relation entre  $\delta_{\text{NL}}$  et  $\delta_{\text{L}}$  pour différentes cosmologies et/ou modèles : la théorie linéaire (vert), le modèle numérique de l’effondrement sphérique pour un Univers Einstein-de-Sitter ( $\Omega_{\text{m}} = 1$ ,  $\Omega_{\Lambda} = 0$ ,  $\Omega_K = 0$ ) (noir),  $\Lambda$ CDM ( $\Omega_{\text{m}} = 0.30$ ) (rouge), la relation de Bernardeau (1994) (bleu), le meilleur ajustement (3.26) pour une cosmologie  $\Lambda$ CDM (magenta).

Le fait remarquable est que les paramètres cosmologiques n’interviennent plus ici. Néanmoins, il ne s’agit que d’une approximation. Si on utilise cette équation en conjonction avec (3.22), on trouve la relation suivante :

$$\delta_{\text{L}} = -\frac{3}{2} \left[ \left( 1 - \frac{1}{3} \delta_{\text{NL}} \right)^2 - 1 \right] \quad (3.24)$$

On voudrait vérifier que cette formule donne un résultat précis pour différentes cosmologies en utilisant le modèle numérique développé dans la section précédente. Pour cela on utilise la procédure suivante (on travaille à  $a_0$  fixé, cf. équation (3.17)) :

- on choisit un  $\delta_{\text{L}}$
- on utilise la théorie linéaire pour trouver la condition initiale qui correspond pour le collapse sphérique

$$\delta_0 = \frac{\delta_{\text{NL}} D(a=1)}{D(a_0)} . \quad (3.25)$$

- on utilise le modèle sphérique de l’équation (3.11) pour trouver le rayon final de la surdensité  $b(a=1)$ , ou de manière équivalente la fluctuation  $\delta_{\text{f}}$ .
- on peut finalement calculer  $\delta_{\text{NL}}$  à l’aide de la relation (3.22).

On a représenté sur la figure 3.15 la relation numérique qui a été trouvée entre  $\delta_{\text{L}}$  et  $\delta_{\text{NL}}$  pour différent modèle d’Univers et pour la formule de Bernardeau (1994). On a aussi représenté le résultat donné par un meilleur ajustement sur la courbe donné par le modèle numérique :

$$\delta_{\text{L}} \simeq -1.695 \left[ (1 - 0.285 \delta_{\text{NL}})^2 - 1 \right] . \quad (3.26)$$

Cet ajustement donne une correction de l’ordre de quelques pourcents dans les régimes assez non-linéaires, c’est-à-dire pour  $\delta_{\text{L}} \gtrsim 1$ . On peut aussi remarquer qu’elle a une forme très proche de celle donnée grâce à la relation de Bernardeau (1994), à l’équation (3.24).

### 3.5.3 Application au champ de déplacement MAK

On peut maintenant essayer d’appliquer la relation (3.26) pixel par pixel au champ de densité  $\delta_{\text{NL}}$  reconstruit par MAK pour la simulation et FullMock. Le résultat de la transformation est donné par

le panneau de droite de la figure 3.13. On voit que l'on arrive en effet à corriger le biais du côté des  $\delta < 0$ . Néanmoins du côté  $\delta > 0$  cette correction ne permet pas d'améliorer les choses. Ce point n'est pas encore totalement éclairci. On a néanmoins présenté, figure 3.16, les distributions de densité de probabilité d'avoir une fluctuation  $\delta$  sur le volume de la simulation pour une reconstruction effectuée avec les particules de la simulation. On voit que, à l'échelle de lissage de  $12 h^{-1}\text{Mpc}$  avec un filtre top-hat, l'accord entre la PDF prédite par MAK (courbe en trait interrompu) et celle donnée par les conditions initiales (ligne solide) de la simulation est remarquable dans les panneaux du haut. Elle l'est d'autant plus quand on compare à ce que donne la théorie eulérienne linéaire sur le champ de densité (courbe en pointillé-trait interrompu). On note que la PDF non corrigée n'a pas assez d'amplitude dans les vides. Cependant, après correction (panneau du milieu), il est difficile de séparer les deux courbes (originale et reconstruite par MAK) dans les parties  $\delta < 0$  et ceci sur *quatre* ordres de grandeur de l'amplitude de la PDF (diagramme semi-logarithmique à droite). Par contre, il reste un excès de probabilité à  $|\delta| \lesssim 0.1$ . On peut noter aussi l'effondrement de la PDF corrigée aux environs de  $\delta \simeq 1.6$  : cela est dû à ce que la densité linéaire maximale que l'effondrement sphérique est capable de décrire est proche de  $5/3$  (Peebles, 1980).

La figure 3.17 donne ces mêmes courbes pour une reconstruction  $128^3$  sur le catalogue FullMock. Rien ne change quasiment, excepté que l'on a un excès de probabilité pour des  $\delta > 1.6$  pour la PDF non corrigée, ce qui ne change rien au résultat par rapport à la figure précédente quand on applique la correction au champ de densité.

On peut se demander si l'utilisation de la formule (3.26) à la place de la formule de Bernardeau (1994) apporte une vraie valeur ajoutée à la correction du champ de densité. Si on regarde la figure 3.18, on constate que cela n'apporte rien pour  $\delta > -1$ . Par contre, si on souhaite tester la forme de la distribution dans les régions vides ( $\delta < -1$ ) la formule (3.26) doit absolument être utilisée car elle n'introduit pas de biais.

## 3.6 Conclusion

Cette dernière étude sur la forme de la PDF conclue les tests de l'efficacité de la reconstruction MAK sur des distributions de matière noire dont on connaît tout. Il semble que la reconstruction MAK soit quasiment parfaite aux grandes échelles ( $\gtrsim 5 - 10 h^{-1}\text{Mpc}$ ). On va à présent se pencher sur un certain nombres de biais observationnels qui peuvent nous poser problème pour construire le maillage de MAK.



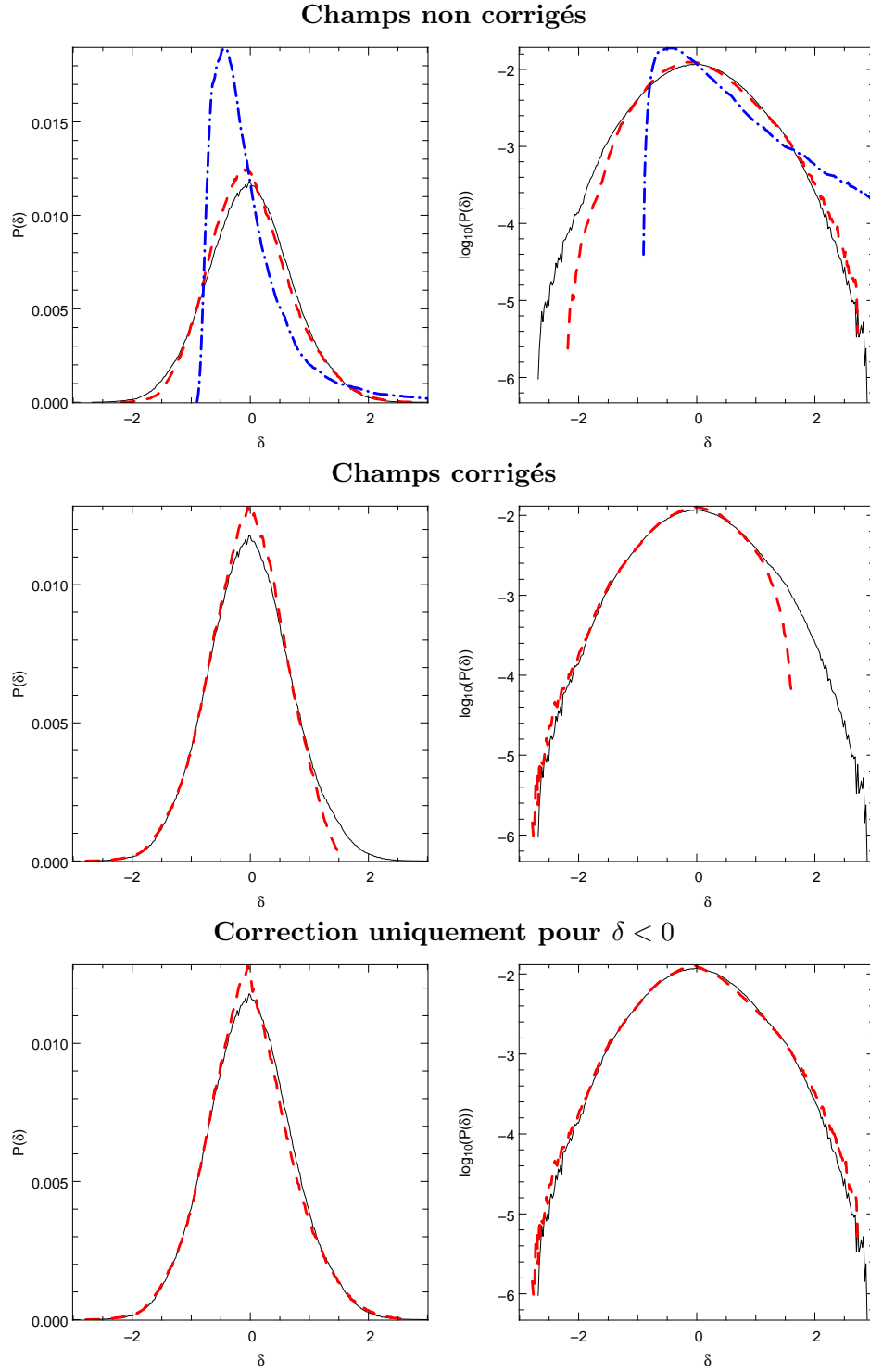


FIG. 3.16: *Reconstruction de la densité de probabilité de fluctuations sur la simulation* – Illustration de l’estimation de la densité probabilité des fluctuations (PDF) du champ de densité initial de la simulation. Les champs ont tous été lissés avec un top-hat de  $12 h^{-1}\text{Mpc}$  de rayon. Les PDF ont été normalisées indépendamment. On a représenté en noir (ligne solide) la PDF des conditions initiales elle-même, en rouge (ligne en trait interrompu) la PDF des conditions initiales reconstruites par MAK, en bleu (ligne en trait interrompu/point) la PDF des conditions initiales reconstruites en utilisant la théorie linéaire sur les conditions finales. La série de panneau de gauche donne ces PDF en échelle linéaire, tandis qu’à droite on a représenté le logarithme décimal de ces PDFs. Les panneaux du haut donnent ces trois PDF pour les champs non corrigés de l’effondrement sphérique. Les panneaux du milieu ont été obtenu en corrigeant le champ de densité donné par MAK+Zel’dovich de l’effondrement sphérique. Les panneaux du bas montrent ce qui se passe si on ne corrige de l’effondrement que pour les  $\delta < 0$ .

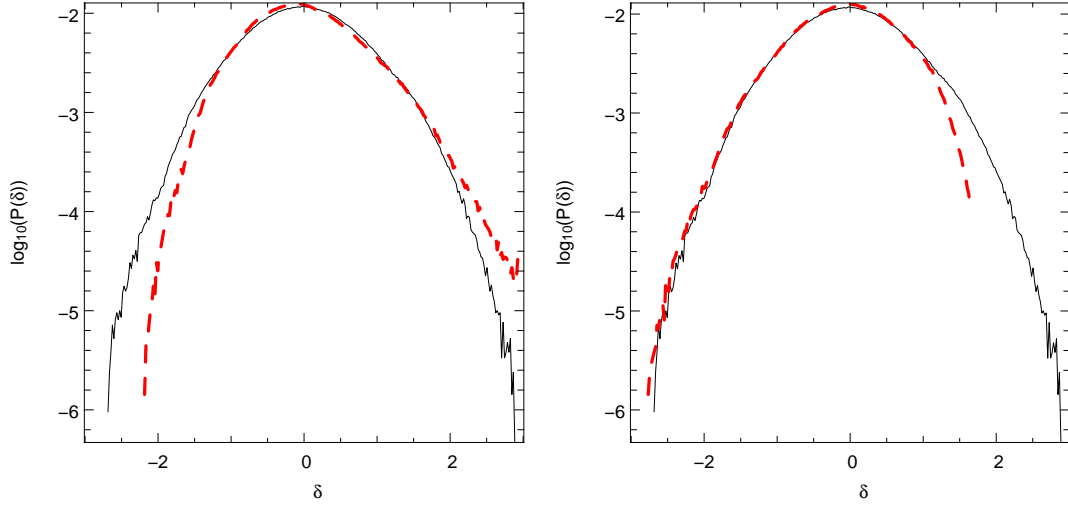


FIG. 3.17: *Reconstruction de la densité de probabilité de fluctuations sur FullMock* – Même densités de probabilité que sur la figure 3.16 mais en utilisant le catalogue *FullMock* (reconstruction  $128^3$ ).

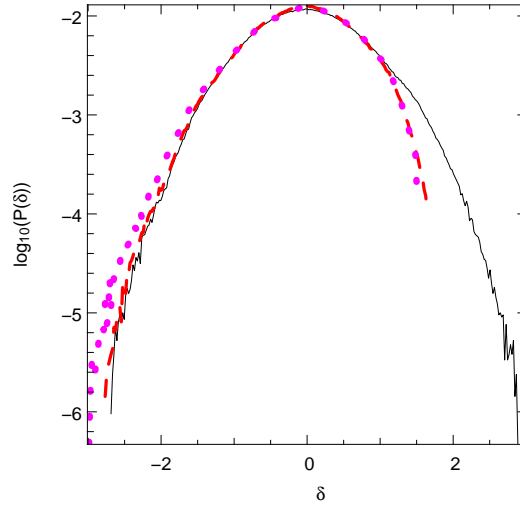


FIG. 3.18: *Reconstruction de la densité de probabilité de fluctuations sur FullMock* – On fait ici de même que dans la figure 3.17 mais on compare le résultat donné par la correction issu de la formule de Bernardeau (1994) (ligne pointillé magenta) ou en utilisant la formule (3.26).



# 4 Analyse bayésienne des champs de vitesse

## Sommaire

<b>4.1 Les erreurs dans la reconstruction MAK</b>	<b>51</b>
4.1.1 Distribution des erreurs	51
4.1.2 Outils pour quantifier les erreurs	52
<b>4.2 Mesurer <math>\Omega_m</math></b>	<b>55</b>
4.2.1 La méthode des moments	55
4.2.2 La méthode de la fonction de vraisemblance	55
<b>4.3 Analyse bayésienne</b>	<b>56</b>
4.3.1 Le modèle bayésien	56
4.3.2 Méthode numérique de calcul de la fonction de vraisemblance	61
4.3.3 Test de la méthode sur des données générées aléatoirement	62
4.3.4 Analyse bayésienne : introduction de la corrélation	63

**M**AINTENANT que nous avons vu que la méthode de reconstruction MAK est à même de retrouver le champs de vitesse, on souhaiterait pouvoir l’exploiter pour effectuer des mesures de paramètres cosmologiques. Le paramètre qui apparaît le plus simplement dans la reconstruction est sous la forme du coefficient de proportionnalité  $\beta$  [cf. équation (2.75)], qui dépend essentiellement de  $\Omega_m$  par la relation, dans un Univers  $\Lambda$ CDM,  $\beta = \Omega_m^{5/9}$ . On va à présent chercher à retrouver la valeur de ce coefficient en utilisant les vitesses “observées” (issu de la simulation dans un premier temps) comparées aux déplacements reconstruits. Si la stratégie statistique est bonne, on doit être capable d’obtenir le  $\Omega_m$  de l’Univers. Maintenant, il reste à bâtir un modèle statistique capable de pouvoir donner une mesure significative avec un nombre limité voire très limité de points de mesure (typiquement pour les observations on a un millier de mesures de vitesse radiale). Pour cela nous allons d’abord étudier la statistique des erreurs sur la reconstruction des vitesses par la méthode MAK (section 4.1), puis nous verrons deux approches pour estimer le paramètre  $\beta$  (et donc  $\Omega_m$ , section 4.2), afin nous détaillerons l’approche bayésienne de la mesure de  $\Omega_m$  (section 4.3).

## 4.1 Les erreurs dans la reconstruction MAK

### 4.1.1 Distribution des erreurs

Avant d’aller dans les détails des problèmes observationnels, nous allons essayer de clarifier les erreurs intrinsèques à la reconstruction MAK. Premièrement, il y a une dispersion, qu’on attend petite (Mohayaee et al., 2006), dans la reconstruction du champ de déplacement lui-même. Deuxièmement, il y a une dispersion due à l’approximation de Zel’dovich qu’on utilise pour convertir le champ de déplacement en champ de vitesse et pour corriger des distortions de redshift. Une connaissance précise de la distribution des erreurs sur les vitesses reconstruites est en fait requise pour une analyse en fonction de vraisemblance qu’on introduira dans la section 4.3.1. Dans cette section, nous mesurons une telle distribution dans l’espace réel alors que son pendant dans l’espace redshift sera étudié dans la section 5.3. En principe, on s’attend à ce que la largeur de cette distribution augmente quand les problèmes observationnels sont pris en compte alors que sa forme ne devrait pas changer significativement. Nous considérons, dans cette section, des reconstructions basées sur le catalogue *FullMock*, pour lequel les conditions de bord périodiques sont appliquées pour minimiser les problèmes d’effet de bord. Nous supposons aussi que nous connaissons la masse des objets décrits par le catalogue

(halos et particules individuelles). Les reconstructions suivantes ont une résolution comprise entre  $64^3$  et  $128^3$  élément pour le maillage. Nous présenterons donc deux reconstructions obtenus avec deux maillage de MAK différents,  $128^3$  and  $64^3$ , obtenus en utilisant la procédure présentée à la section 3.3.2. Les résultats des champs de déplacement reconstruits sont donnés par la figure 4.1. Ces graphiques donnent la distribution des différences,  $P_{DE}$ , entre la composante sur la ligne de visée du déplacement reconstruit et du “vrai”, donné par la simulation.

Les lignes en pointillés et en tirets correspondent à un ajustement des moindres carrés à la fonction  $P_{DE}$  correspondant à une reconstruction  $128^3$  respectivement avec une fonction gaussienne et lorentzienne donnée par

$$P_{Lor}(x) = \frac{1}{\pi B} \frac{1}{1 + \frac{x^2}{B^2}}. \quad (4.1)$$

L’examen de la figure 4.1 incite au choix de l’approximation lorentzienne avec  $B = 35 \text{ km s}^{-1}$ , qui reproduit mieux que la gaussienne les longues queues de  $P_{DE}$ .

La largeur,  $B$ , de  $P_{DE}$  est assez petite comparée à la dispersion selon la ligne de visée  $\langle \beta^2 \Psi_r^2 \rangle^{1/2} \simeq 292 \text{ km s}^{-1}$ , comme on s’y attendait. Naturellement, la fonction  $P_{DE}$  est légèrement plus plate et plus large pour le cas  $64^3$  que pour le  $128^3$ . Cependant, les extrémités des queues de  $P_{DE}$  sont les mêmes pour les cas  $64^3$  et  $128^3$ . Dans ce régime, les mesures ne sont pas influencées par la résolution de la grille utilisées pour calculer la reconstruction mais plutôt par l’incapacité de MAK à reproduire la dynamique interne des objets massifs et relaxés (Mohayaee et al., 2006).

La figure 4.2 est semblable à la figure 4.1 mais considère la composante sur la ligne de visée des vitesses reconstruites vs les vitesses “exactes”. Bien que l’approximation de Zel’dovich introduise un bruit additionnel, comme cela est montré par une largeur plus grande de la distribution,  $P_{DE}$  reste grossièrement lorentzienne avec une largeur faible  $B = 48 \text{ km s}^{-1}$ . Cette variance est approximativement 50% plus petite que la variance du champ de vitesse sur le volume de la simulation (section 3.1.2). Nous sous-estimons donc légèrement les effets de variance cosmique qui auraient pu être induit par les modes plus grand que la taille de la boîte de la simulation. Néanmoins cela ne changera pas l’essentiel de nos résultats en espace réel, et nous verrons qu’en espace redshift la largeur  $B$  augmente d’un peu moins d’un facteur deux, ce qui l’amène au même niveau que la variance du champ de vitesse.

#### 4.1.2 Outils pour quantifier les erreurs

Afin de vérifier la précision des reconstructions, nous calculons les moments de la distribution de probabilité jointe des vitesses reconstruites  $\mathbf{v}_{rec,i}$  d’un objet  $i$  et de la vitesse simulée de cet objet  $\mathbf{v}_{sim,i}$ . Nous écrivons  $\langle A \rangle$  pour l’opération de moyennage de la quantité  $A$

$$\langle A \rangle = \frac{1}{N} \sum_{i=1}^N A_i. \quad (4.2)$$

Nous définissons les trois moments d’ordre deux :

$$\sigma_r^2 = \langle v_{rec}^2 \rangle - \langle v_{rec} \rangle^2, \quad \sigma_s^2 = \langle v_{sim}^2 \rangle - \langle v_{sim} \rangle^2, \quad \sigma_{r,s} = \langle (v_{rec} - \langle v_{rec} \rangle) (v_{sim} - \langle v_{sim} \rangle) \rangle \quad (4.3)$$

A partir de ces moments, on peut construire le coefficient de corrélation

$$r = \frac{\sigma_{r,s}}{\sigma_r \sigma_s} \quad (4.4)$$

et le rapport entre la largeur de la PDF du champ de vitesse reconstruit et la largeur de la PDF du champ simulé :

$$s = \frac{\sigma_r}{\sigma_s}. \quad (4.5)$$

Pour ces deux quantités, la valeur optimale est 1. Alternativement, deux autres estimateurs de la “pente” des vitesses reconstruites versus celles simulées peuvent être construits à partir des moments ci-dessus :

$$s_{min} = \frac{\sigma_{r,s}}{\sigma_s^2} = sr \quad \text{and} \quad s_{max} = \frac{\sigma_r^2}{\sigma_{r,s}} = s/r. \quad (4.6)$$

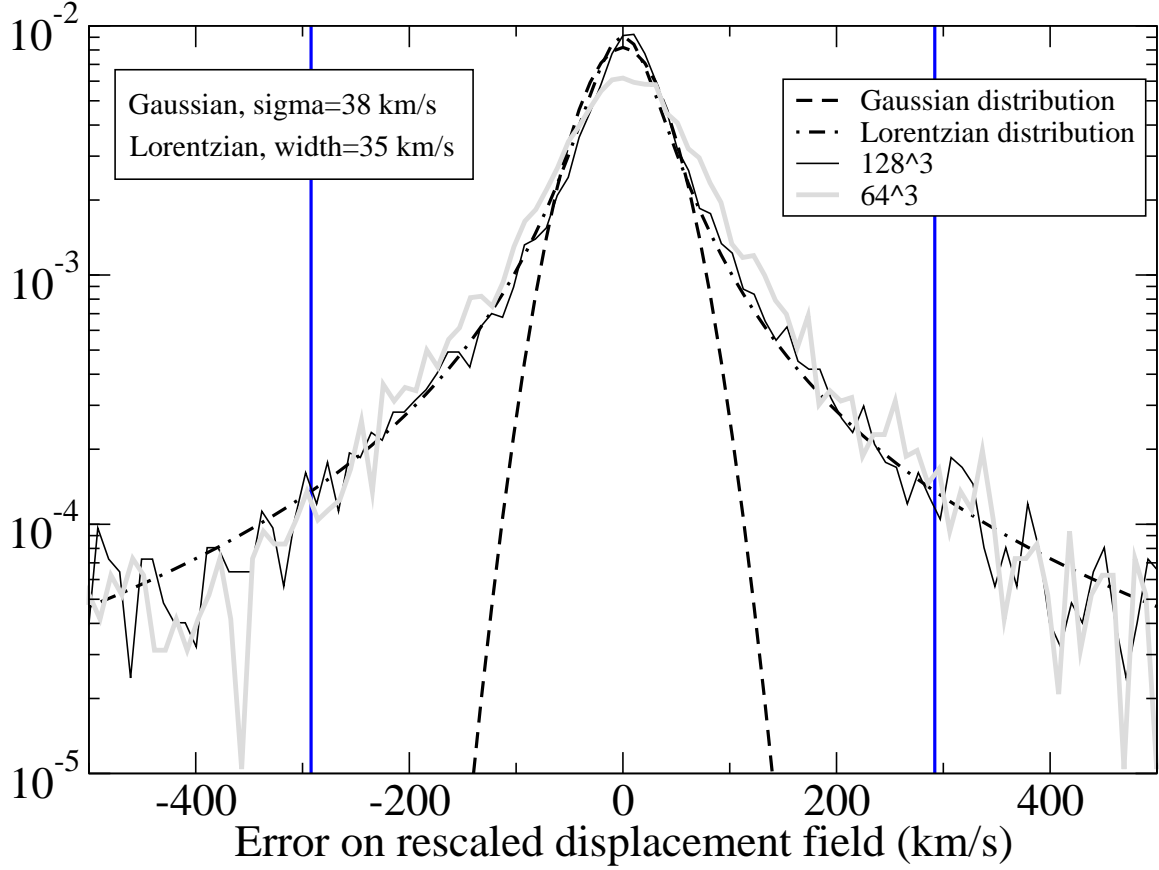


FIG. 4.1: *Statistique des erreurs dans les déplacements reconstruits* – Ce graphique montre la distribution de probabilité de la quantité  $\beta(\Psi_{r,rec} - \Psi_{r,sim})$  mesurée dans *FullMock* (ligne solide), où  $\Psi_{r,rec}$  et  $\Psi_{r,sim}$  sont les composantes respectives des champs de déplacements reconstruits et simulés, après avoir choisi un observateur au centre de la boîte de simulation. Les lignes en tirets et point-tirets donne respectivement le meilleur ajustement à une distribution gaussienne et lorentzienne.

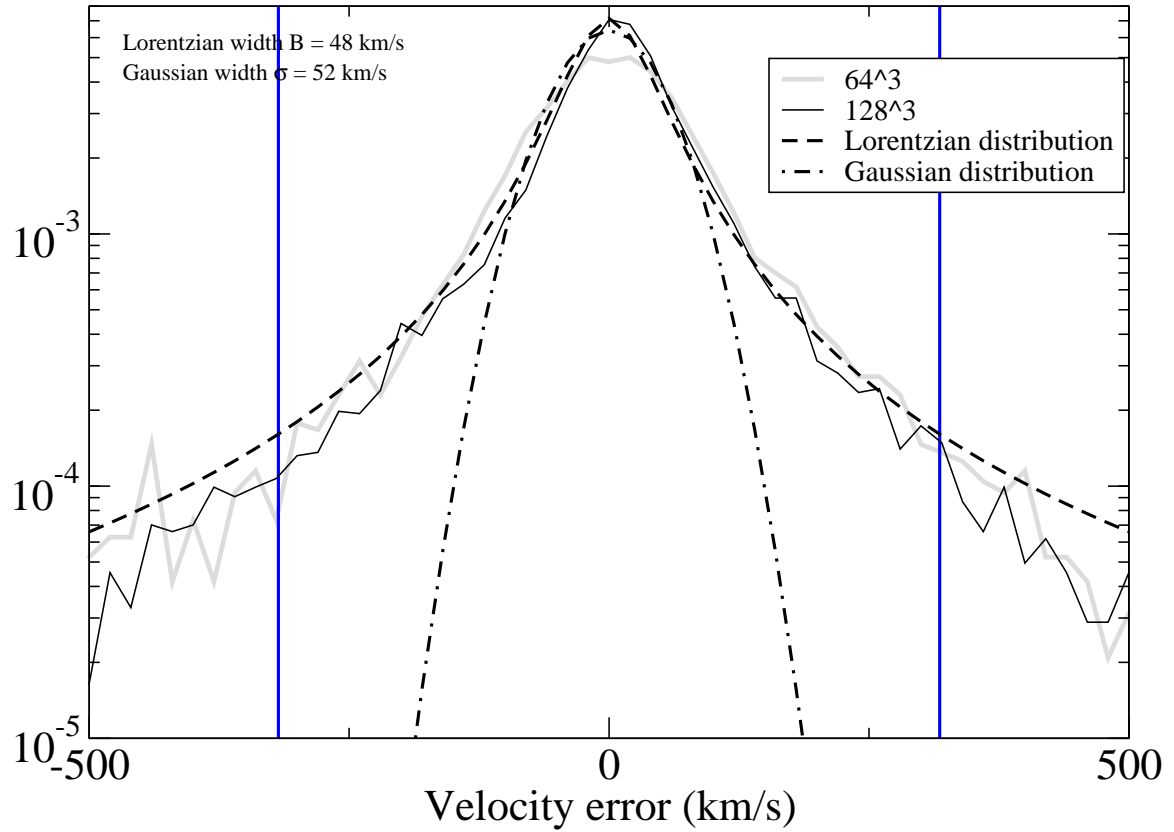


FIG. 4.2: *Statistique des erreurs des vitesses reconstruites* – Même chose que pour la figure 4.1 mais la ligne solide correspond à la distribution de probabilité de la quantité  $v_{r,rec} - v_{r,sim}$ , où  $v_{r,rec}$  et  $v_{r,sim}$  sont respectivement les composantes selon la ligne de visée des vitesses reconstruites et simulées.

Ces deux pentes sont intéressantes quand on fait une estimation de  $\Omega_m$  à partir de  $s$  et que l'on a besoin d'une évaluation de l'incertitude. Les deux pentes supplémentaires déterminées de cette façon devrait, idéalement, être égale à  $s$  mais à cause d'un manque d'une corrélation parfaite ( $r < 1$ ), elles sont en fait différentes dans les cas réalistes. En fait, nous avons  $s_{\min} < s_{\text{med}} < s_{\max}$ .

Il faut par ailleurs remarquer que nous pouvons définir la dispersion relative

$$\sigma^2 = \frac{\langle (v_{\text{rec}} - v_{\text{sim}})^2 \rangle}{\sigma_s^2} = 1 + s^2 - 2sr, \quad (4.7)$$

qui est une mesure du rapport bruit-sur-signal : un grand  $\sigma$  correspond à un signal bas. Idéalement, on voudrait  $\sigma = 0$ .

## 4.2 Mesurer $\Omega_m$

A l'aide de la relation entre les déplacements reconstruits et les vitesses observées, il est possible d'avoir un estimateur de  $\Omega_m$ . Nous allons présenter deux méthodes de niveau de difficultés différents et présentant des avantages et des inconvénients différents.

### 4.2.1 La méthode des moments

La première méthode consiste à fabriquer une fonction de distribution de probabilité entre les vitesses “observées” et les vitesses reconstruites ( $P(v_{\text{sim}}, \psi_{\text{rec}})$ ). Pour faire en sorte de se focaliser sur les déviations par rapport à la distribution attendue nous ne représenterons pas  $P(v_{\text{sim}}, \psi_{\text{rec}})$ , mais  $P(v_{\text{sim}}, v_{\text{rec}})$ , c'est-à-dire la distribution jointe des vitesses reconstruites et simulées, avec  $v_{\text{rec}} = \Omega^{5/9} \psi_{\text{rec}}$ . En principe, le calcul des moments (section 4.1.2) de cette distribution devrait permettre d'obtenir une pente de 1.

Néanmoins, comme nous l'avons vu à la section 3.4, les queues non-linéaires donnent une forme d'hélice à  $P(v_{\text{sim}}, v_{\text{rec}})$ , ce qui est susceptible d'introduire un petit biais sur la comparaison vitesse-vitesse. Par exemple, on peut estimer la pente du diagramme de dispersion en bas à gauche de la figure 3.9 en utilisant le rapport  $s = \sigma_{v,\text{rec}} / \sigma_{v,\text{sim}}$ , où  $\sigma_{v,\text{rec}}^2$  et  $\sigma_{v,\text{sim}}^2$  sont respectivement les variances des champs de vitesses reconstruits et simulés. Dans ce cas, le paramètre  $\beta$  est biaisé vers des valeurs plus élevées d'à peu près 7%. Cependant, une inspection visuelle du diagramme montre qu'il ne devrait pas y avoir de biais si seulement la partie centrale du nuage était utilisée pour le calcul de la pente. Pour y arriver, nous allons d'abord appliquer un filtre adaptatif SPH (annexe E) au nuage de points pour produire une fonction de densité de probabilité (PDF), qui est échantillonnée par les points du nuage, sur une grille régulière. Nous pouvons maintenant calculer l'isocontour à  $1.5\sigma$ , qui enclot la région où l'intégrale de la PDF est égale à 68%. Cette méthode a déjà été utilisée avec succès par Colombi et al. (2007) pour la comparaison gravité-vitesse. Seuls les points à l'intérieur du contour à  $1.5\sigma$  sont utilisés pour calculer le nouveau coefficient  $s_{\text{med},68}$ . Le paramètre  $\beta$  déduit de  $s_{\text{med},68}$  est maintenant statistiquement non-biaisé. De manière semblable, nous définissons deux autres estimateurs de la pente  $s_{\min,68}$  et  $s_{\max,68}$ , comme il a été discuté dans la section 4.1.2. Dans ce travail, jusqu'à la section 5.5, nous ne discuterons que les mesures de  $\Omega_m$  obtenues à travers l'estimateur  $s_{\text{med},68}$ . Le  $\Omega_m$  obtenu par cette méthode est identifiée par un “ $1.5\sigma$ ” pour faire la différence avec celui obtenu par l'analyse en fonction de vraisemblance que nous allons établir dans la section suivante et qui est identifié par un “ $\mathcal{L}$ ” dans les tables et figures. Nous verrons un test de cette méthode sur des nuages de points simulés, dont la forme est construite sur la base de l'analyse sur les erreurs de reconstruction, dans la section 4.3.3.

### 4.2.2 La méthode de la fonction de vraisemblance

Une méthode très utilisée à présent en cosmologie observationnelle utilise une technique basée sur la maximisation d'une probabilité. Cette technique permet d'extraire en principe toute l'information



contenue dans un jeu de données. Le problème est de définir la probabilité à maximiser. Usuellement on utilise des gaussiennes, le problème c’est que des lois de probabilités gaussiennes pour les données ne sont valables que dans la limite des grands nombres (cf. théorème limit-central). En cosmologie observationnelle on est exactement dans le cas où le nombre de données est plutôt faible. Dans notre cas, on est proche de quelques milliers de points de mesure pour les distances des galaxies par exemple. Nous allons donc montrer dans la section suivante comment décrire d’une manière plus quantitative une comparaison statistique des vitesses reconstruites avec les vitesses observées afin d’obtenir une maximisation non biaisée de  $\Omega_m$ .

## 4.3 Analyse bayésienne

Une première étape dans cette analyse statistique est de considérer que les vitesses observées sont non-corrélées. Nous allons donc supposer dans cette partie que les points de mesure sont tous statistiquement *indépendants*.

### 4.3.1 Le modèle bayésien

L’observation des galaxies nous donne d’abord accès à leur distance et non leur vitesse propre radiale. Une méthode basée sur les distances pour faire une comparaison entre modèles et observations est potentiellement moins sensible aux erreurs sur les distances. En effet, si on compare directement les distances, on doit faire face à une petite erreur relative sur chaque mesure au lieu d’une erreur immense si on considère les vitesses propres. Nous allons ici discuter les problèmes liés à la manière de sélectionner les galaxies et la calibration du point-zéro dans les mesure de distances. Les notations utilisées sont celles de Strauss & Willick (1995).

#### Présentation de la chaîne Bayésienne

Dans la relation de Tully-Fisher (TF), on estime la magnitude absolue d’une galaxie comme une fonction de la largeur de la raie HI : la pente entre ces deux quantités peut être biaisée car l’échantillon considéré de galaxies a été sélectionné selon la magnitude apparente (Strauss & Willick, 1995). Cet effet, connu sous le nom *de biais de sélection*, est purement statistique et, s’il n’est pas pris correctement en compte, peut mener à d’énormes erreurs systématiques. En utilisant les magnitudes absolues des galaxies dérivées grâce à cette relation et la magnitude apparente de ces mêmes galaxies, on peut former le *module de distance* :

$$\mu(r) = m(r) - M = 5 \log_{10} \left( \frac{r}{10 \text{ pc}} \right) \quad (4.8)$$

avec  $r$  la distance de la galaxie considérée. Pour construire cette quantité dans le cas de galaxies groupées, processus que l’on sera amené à utiliser, il convient d’employer la magnitude absolue et celle apparente du même groupe de galaxies. On parlera en général d’objets d’un catalogue pour désigner une galaxie ou un groupe de galaxies. En plus des biais statistiques dont on vient de parler, les vitesses propres obtenues par une reconstruction Lagrangienne, ici par exemple MAK, sont parfois très bruitées, comme le montre la figure 5.19 plus loin. Un effet plus subtil est introduit par la distribution gaussienne des vitesses de notre échantillon que nous allons analyser. Nous devons tenir compte de cet “effet de sélection” afin d’éviter d’être contaminé par d’éventuelle grosses erreurs dans la reconstruction, en particulier pour des objets avec de grandes vitesses propres. Nous avons donc besoin d’une approche Bayésienne pour tenir compte de tous ces effets statistiques.

En principe, la fonction de vraisemblance donne la probabilité d’occurrence pour les données, c’est-à-dire ici les positions redshift  $\mathbf{z} = \{\mathbf{z}_i\}$ , avec  $i$  allant de 1 à  $N$ , et les modules de distance  $\mathbf{M} = \{\mu_i\}$ , en supposant un modèle cosmologique décrit par le vecteur de paramètre  $p$ . De plus, nous supposons que nous avons une estimation des erreurs de mesure sur  $\mathbf{M}$  par l’intermédiaire de l’ensemble  $\mathbf{S}$ . La description précise de  $\mathbf{S}$  sera donnée dans le prochain paragraphe. Typiquement les erreurs sur les mesures de redshift sont de l’ordre de 50-60 km s<sup>-1</sup>. Cela signifie que l’on peut les estimer comme

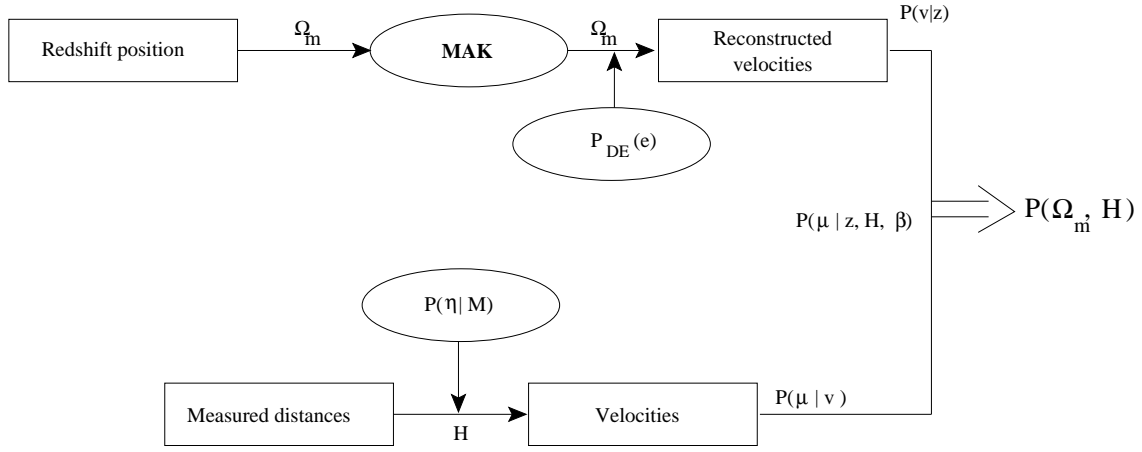


FIG. 4.3: *Analyse en maximum de la fonction de vraisemblance* – Ce dessin illustre le principe de la chaîne bayésienne utilisée pour établir la fonction de vraisemblance. Les données d’entrée sont localisées sur la gauche et la distribution postérieure  $P(\beta, H)$  en sortie à droite.

négligeables si on considère des objets plus loins que  $R_z = 6 - 10$  Mpc/h. Le champ de vitesse dans ce volume est, dans tous les cas, aussi mal reconstruit à cause de la singularité introduite par les distortions dans l’espace redshift par les vitesses propres, dans une région proche de l’observateur (§ 5.3). Dans l’analyse qui suit, nous considérerons que les erreurs sur les mesures en redshift comme négligeables en évitant les objets situés à moins de  $10 h^{-1}$  Mpc de l’observateur, ainsi nous avons : <sup>1</sup>

$$P(\mathfrak{M}, \mathfrak{Z}|p, \mathfrak{S}) \propto P(\mathfrak{M}|\mathfrak{Z}, p, \mathfrak{S}) = \mathcal{L}(p) \quad (4.9)$$

La fin de cette section est dévolue à calculer le membre de droite de cette équation. Pour y arriver, nous allons décomposer cette probabilité en petits morceaux :

$$P(\mathfrak{M}|\mathfrak{Z}, \mathfrak{S}, p) = \iiint_{\mathfrak{M}_r, \mathcal{V}, D} P(D) \times P(\mathfrak{M}|\mathfrak{M}_r, \mathfrak{S}, D, p) \times P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p) \times P(\mathcal{V}|\mathfrak{Z}, p) d\mathfrak{M}_r d\mathcal{V} dD \quad (4.10)$$

avec  $\mathfrak{M}_r = \{\mu_{1,r}, \dots, \mu_{N,r}\}$  représentant les “vrais” modules de distance,  $\mu_{i,R} \in [-\infty, +\infty]$  et  $\mathcal{V} = \{v_1, \dots, v_n\}$  les “vrais” vitesses propres des objets  $P(\mathfrak{M}|\mathfrak{M}_r, \mathfrak{S}, p)$  est la probabilité de mesurer l’ensemble des modules de distance  $\mathfrak{M}$  sachant qu’on a l’ensemble des vrais modules de distance  $\mathfrak{M}_r$  et que l’erreur attendue sur la mesure est donnée par  $\mathfrak{S}$ .  $P(\mathfrak{M}_r|\mathcal{V}, p)$  est la probabilité d’obtenir l’ensemble de modules de distance  $\mathfrak{M}_r$  sachant que les vitesses reconstruites sont données par  $\mathcal{V}$ .  $P(\mathcal{V}|\mathfrak{Z}, p)$  est la probabilité que les vitesses sont bien reconstruites à partir des données redshift  $\mathfrak{Z}$ . La probabilité  $P(D)$  sera introduite dans le dernier paragraphe pour tenir compte d’une incertitude dans la calibration de la relation de Tully-Fisher. Toutes ces probabilités sont calculées en supposant le vecteur de paramètres du modèle  $p$ . Nous allons établir la fonction de vraisemblance  $\mathcal{L}(p)$  en trois étapes :

- D’abord, les distributions d’erreur liées à l’incertitude sur les observations sont prises en compte pour obtenir un estimateur non-biaisé pour les groupes de galaxies. Cette analyse va nous donner la probabilité  $P(\mu_i|\mu, \sigma_{\mu,i}, p)$ .
- Ensuite, les erreurs sur les vitesses reconstruites sont considérés pour calculer  $P(v|\mathfrak{Z}, p)$ .
- Enfin, ces deux analyses sont fusionnées comme indiqué ci-dessus pour engendrer la fonction de vraisemblance qui donne la distribution postérieure sur  $\beta$  et la constante de Hubble  $H$ .

Un schéma de la chaîne Bayésienne ci-dessus est donné figure 4.3.

<sup>1</sup>Bien que cela soit possible en théorie d’éviter cette hypothèse, il est en pratique très difficile de l’éviter pour des raisons de temps de calcul numérique. En effet, on aurait alors besoin d’exécuter de nombreuses reconstructions MAK afin d’évaluer l’intégrale supplémentaire nécessaire dans l’équation (4.9).

### Distribution d'erreur sur le module de distance

Afin d'établir la fonction de vraisemblance comparant la distance mesurée au champ de vitesse reconstruit, nous supposons que les catalogues de distance sont obtenus en utilisant la relation de Tully-Fisher inverse (Shaya et al., 1995),

$$\eta^0(M) = -e(M + D) , \quad (4.11)$$

où  $M$  est la magnitude absolue de la galaxie considérée,  $\eta_0(M)$  est sa largeur de raie prédite,  $e$  est la pente, et  $D$  est la calibration de point zéro (les deux dernières valeurs sont supposées connues exactement ici). Il est connu que la relation de Tully-Fisher inverse est moins sensible aux effets de sélection que la relation TF directe (Strauss & Willick, 1995). Les données observationnelles montrent que les différences entre les largeurs de raie prédites  $\eta^0(M)$  et celles mesurées  $\eta$  pour une galaxie de magnitude absolue  $M$  sont distribuées de manière Gaussienne<sup>2</sup> (Pizagno et al., 2006; Tully & Pierce, 2000). Par conséquent, la probabilité de mesurer une largeur de raie  $\eta$ , sachant que la galaxie a une magnitude absolue  $M$ , et supposant que la relation de Tully-Fisher est connue, est :

$$P(\eta|M, e, D) = \frac{1}{\sqrt{2\pi}\sigma_\eta(M)} e^{-\frac{(\eta - \eta^0(M))^2}{\sigma_\eta^2(M)}} \quad (4.12)$$

avec  $\sigma_\eta(M)$  l'erreur estimée sur la largeur de raie pour une magnitude absolue  $M$ . Les catalogues de distance sont composés de modules de distances estimés  $\mu_e$  à partir de la relation de TF inverse. Ces modules estimés sont construits à partir de la statistique d'un seul groupe. Par conséquent, la probabilité jointe d'avoir une galaxie dans un groupe avec à la fois une largeur de raie  $\eta$  et une magnitude absolue  $M$ , en supposant la relation de TF  $\eta^0(M)$ , est :

$$P(\eta, M|e, D) = F(M)P(\eta|M, e, D) , \quad (4.13)$$

avec  $F(M)$  la fonction normalisée de luminosité absolue pour un groupe<sup>3</sup>

L'estimation du module de distance est donnée par :

$$\mu_e = m - M_e(\eta) = M + \mu_0(r) + D' + \frac{\eta}{e'} , \quad (4.14)$$

où  $D'$  et  $e'$  sont les paramètres estimés de la relation inverse de TF de l'équation Eq. (4.11) et  $\mu_0(r) = 5 \log(\frac{r}{10 \text{ pc}})$  le vrai module de distance du groupe considéré. La probabilité conditionnelle que le module de distance estimé pour le groupe est  $\mu$  peut être écrite selon l'expression suivante, en supposant que les paramètres de Tully-Fisher sont  $e'$  et  $D'$  et que les vrais paramètres pour ce groupe sont  $e$  et  $D$  :

$$\begin{aligned} P(\mu|\mu_0(r), e, e', D, D') &= \langle \delta_D(\mu - \mu_e) \rangle_{\text{group}} = \left\langle \delta_D \left( \mu - \mu_0(r) + M + D' + \frac{\eta}{e'} \right) \right\rangle_{\text{group}} \\ &= \int_M \frac{e' F(M)}{\sqrt{2\pi}\sigma_\eta(M)} e^{-\frac{(e'(\mu_0(r) - \mu) + e' D' - e D + (e' - e) M)^2}{2\sigma_\eta^2(M)}} dM . \end{aligned} \quad (4.15)$$

Quand on utilise la relation inverse de TF, on suppose que la pente  $e'$  est complètement déterminée et que  $e' = e$ . Comme l'erreur  $\sigma_\eta(M)$  varie peu avec  $M$ , l'approximation  $\sigma_\eta(M) \simeq \sigma_0$ , avec  $\sigma_0$  constant, peut-être utilisée. La probabilité précédente se réduit à

$$P(\mu|\mu_0(r), e, D, D', \sigma_0) = \frac{e}{\sqrt{2\pi}\sigma_0} e^{-\frac{e^2(\mu - \mu_0(r) + D' - D)^2}{2\sigma_0^2}} . \quad (4.16)$$

<sup>2</sup>En fait, en écrivant l'équation (4.12), deux effets sont mélangés : l'erreur sur la mesure de la largeur de raie, qui peut atteindre 10% à cause de l'incertitude sur la correction de l'inclinaison galactique, et l'erreur intrinsèque de modélisation par la relation de Tully-Fisher.

<sup>3</sup>On peut remarquer ici que la fonction de sélection est supposée indépendante de  $\eta$  et qu'ainsi elle est absorbée dans  $F(M)$ .  $F(M)$  correspond à  $\Phi(M)S(M, \eta)$  dans les notations de Strauss & Willick (1995), par exemple equation (188).

Bien que la pente  $e'$  soit bien déterminée, la calibration de point zero  $D$  peut encore être affectée d'erreurs non négligeables.<sup>4</sup> L'ensemble décrivant les erreurs sur les modules de distance est donc  $\mathfrak{S} = \{\sigma_{0,1}/e, \dots, \sigma_{0,N}/e\} = \{\sigma_{\mu,1}, \dots, \sigma_{\mu,N}\}$ . L'erreur sur cette calibration affectera les distances globalement. Comme première approximation, nous allons modéliser l'erreur sur le point zéro par une gaussienne centrée sur  $D$  avec une erreur standard de  $\sigma_D$ .

*Lien entre module de distance et vitesse propre* – La deuxième fonction de probabilité dans l'équation Eq. (4.10) est  $P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p)$ , qui est en fait une distribution liant les vitesses et les redshifts aux modules de distance. Cela correspond principalement à un changement de variable et nous donnons directement son expression, qui est inspirée de la relation (2.42) :

$$P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p) = \prod_{i=1}^N H10^{\mu_{r,i}/5} \delta_D \left( z_i - v_i - 10 \text{ pc} \times H10^{\mu_{r,i}/5} \right) \quad (4.17)$$

*Distribution des vitesses reconstruites* – Nous allons à présent établir l'expression de  $P(v|\mathfrak{Z}, p)$ . avec le vecteur de paramètres du modèle que nous avons choisi  $p = (H, \beta, B_v, \sigma_v, \gamma_*, e) - \sigma_v$  et  $\gamma_*$  vont être introduit dans les paragraphes qui suivent immédiatement. On peut décomposer  $P(v|\mathfrak{Z}, p)$  de la façon suivante

$$P(v|\mathfrak{Z}, p) = \int_{\mathfrak{P}} P(v|\mathfrak{P}, p) P(\mathfrak{P}|\mathfrak{Z}, p) d\mathfrak{P}, \quad (4.18)$$

avec  $\mathfrak{P} = \{\psi_{r,i}\}$  les déplacements reconstruits. Comme la reconstruction MAK est déterministe une fois qu'on suppose  $\beta$  (§ 5.3), la deuxième distribution de probabilité est simplement donnée dans notre cas par

$$P(\mathfrak{P}|\mathfrak{Z}, p) = \prod_{i=1}^n \delta_D (\psi_{r,i} - \psi_i(\mathfrak{Z}, \beta)) \quad (4.19)$$

avec  $\psi_i$  représentant le déplacement reconstruit par MAK du  $i$ -ème objet, qui est une fonction des coordonnées redshift de tous les objets et de  $\beta$ . Donc, étudier  $P(\mathcal{V}|\mathfrak{Z}, p)$  se réduit à examiner  $P(\mathcal{V}|\mathfrak{P}(p_0), p')$ , avec  $p' = (H, \beta', B_v, \sigma_v, \gamma_*)$ ,  $p_0 = (H, \beta_0, B_v, \sigma_v, \gamma_*)$ ,  $\beta_0$  étant le facteur de croissance supposé pour calculer l'ensemble  $\mathfrak{P}(p_0)$  en utilisant une reconstruction redshift.  $P(\mathcal{V}|\mathfrak{Z}, p)$  et  $P(\mathcal{V}|\mathfrak{P}(p_0), p')$  sont égaux seulement si  $p = p' = p_0$ . Donc on a besoin de multiples reconstructions redshift pour construire la fonction de probabilité  $P(\mathcal{V}|\mathfrak{Z}, p)$ . Travailler avec l'ensemble intermédiaire  $\mathfrak{P}$  étant plus facile qu'avec  $\mathfrak{Z}$ , nous posons la fonction de vraisemblance réduite :

$$\mathfrak{L}'_{\beta_0}(p') = \iint_{\mathfrak{M}_r, \mathcal{V}} P(\mathfrak{M}|\mathfrak{M}_r, \mathfrak{S}) P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p) P(\mathcal{V}|\mathfrak{P}(p_0), p') d\mathfrak{M}_r d\mathcal{V}. \quad (4.20)$$

Nous allons établir l'expression de la fonction de probabilité élémentaire  $P(v_r|\psi_r, p)$  qui nous donnera

$$P(\mathcal{V}|\mathfrak{P}(p_0), p) = \prod_{i=1}^N P(v_{r,i}|\psi_{r,i}, p) \quad (4.21)$$

en supposant l'indépendance statistique de tous les doublets  $\{v_{r,i}, \psi_{r,i}\}$ , et que  $\mathfrak{P}(p_0)$  est obtenu en utilisant une reconstruction redshift pour laquelle  $\beta = \beta_0$ .<sup>5</sup>  $\mathfrak{L}'$  peut être écrite après factorisation :

$$\mathfrak{L}'_{\beta_0}(p') = \prod_{i=1}^N \iint_{\mu_r, v_r} P(\mu_i|\mu_r, \sigma_{0,i}/e) P(\mu_r|v_r, z_i, p) P(v_r|\psi_{r,i}, p) d\mu_r dv_r. \quad (4.22)$$

Le calcul de  $\mathfrak{L}'$  est clairement facilité en utilisant cette forme. Nous pouvons maintenant nous concentrer sur la troisième fonction de probabilité de l'équation ci-dessus.

<sup>4</sup>La dernière calibration est donnée dans Tully et al. (2008).

<sup>5</sup>L'indépendance statistique sera discutée dans la section 4.3.4.

Comme il a été établi dans la section 4.1.1, la distribution des erreurs sur le champ de vitesse reconstruit est la Lorentzienne

$$P_{\text{DE}}(e_{\psi v}) \propto \frac{1}{1 + \left(\frac{e_{\psi v}}{B_v}\right)^2}, \quad (4.23)$$

avec  $B_v = 86 \text{ km s}^{-1}$  (reconstruction redshift), avec  $e_{\psi v}$  la distance entre la vitesse reconstruite  $\beta\Psi_r$  et la vraie vitesse  $v_r$ . Cette formulation est différente par rapport à celle qui consisterait à dire que la vitesse reconstruite est affectée d'une erreur par rapport à la vraie vitesse, et permet quelques erreurs dans le champ de déplacement reconstruit. Inversement, on constate que l'erreur est mieux modélisée de cette manière. Comme il sera vu dans la section 5.4.3, les vitesses reconstruites peuvent aussi avoir un décalage systématique qui doit être enlevé quand on mesure  $\beta$ . L'erreur  $e_{\Psi v}$  est donc

$$e_{\Psi v} = \alpha_* v_r - \beta_* \Psi_r + \gamma_* \quad (4.24)$$

avec

$$\alpha_*^2 + \beta_*^2 = 1 \text{ and } \beta = \beta_*/\alpha_* \quad (4.25)$$

et  $\gamma_*$  pour tenir compte d'un décalage excédentaire dans les vitesses reconstruites. En utilisant la théorie linéaire (Peebles, 1980, ainsi que les section 2.3.2 et C), nous savons que la composante sur ligne de visée du champ de vitesse doit être distribuée suivant une loi gaussienne. Nous supposons maintenant que la probabilité absolue pour un objet d'avoir une vitesse  $v$ , étant donné un modèle cosmologique  $p$ , est donnée par une distribution gaussienne :

$$P_{\text{vel}}(v|p) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{v^2}{2\sigma_v^2}}. \quad (4.26)$$

Il faut remarquer qu'il est vraisemblable que les données observationnelles ne couvrent pas un volume suffisamment grand pour que les vitesses suivent parfaitement cette loi (cf. section 5.4.3). De plus, cet a-priori aura une importance certaine quand on va analyser des données hautement bruitées. Les défauts d'une telle approche seront discutés dans la section 5.5.2 sur des données virtuelles. On peut retrouver un prior uniforme sur les vitesses en prenant la limite  $\sigma_v \rightarrow +\infty$  dans les équations qui suivent. En supposant que la variable aléatoire  $e_{\psi v}$  est indépendante de  $v_r$  et que ces deux quantités sont elle-même statistiquement indépendantes de  $\beta$  et  $\gamma_*$ , nous pouvons maintenant écrire la probabilité jointe de reconstruire  $\psi_r$ , pour un objet ayant une vraie vitesse  $v_r$  :

$$\begin{aligned} P(v_r, \psi_r, \beta, \gamma_* | B_v, \sigma_v) &= \beta_* P_{\text{DE}}(e_{\Psi v} | B_v, \sigma_v) \times P(\beta, \gamma_* | B_v, \sigma_v) \times P(v_r | B_v, \sigma_v) \times P(\psi_r | B_v, \sigma_v) \\ &= \beta_* C(B_v, \sigma_v) P(\beta, \gamma_* | B_v, \sigma_v) \frac{P(\psi_r | B_v, \sigma_v) e^{-\frac{v_r^2}{2\sigma_v^2}}}{1 + \left(\frac{\beta_* \psi_r - \alpha_* v_r + \gamma_*}{B_v}\right)^2}, \end{aligned} \quad (4.27)$$

où  $C$  est une fonction éventuellement dépendante de  $B_v$  et  $\sigma_v$ . La probabilité conditionnelle que la vraie vitesse est  $v_r$  étant donné le déplacement reconstruit  $\psi_r$  est maintenant exactement

$$P(v_r | \psi_r, p) = \frac{e^{-\frac{v_r^2}{2\sigma_v^2}} \left(1 + \left(\frac{\beta_* \psi_r - \alpha_* v_r + \gamma_*}{B_v}\right)^2\right)^{-1}}{\int_{v=-\infty}^{+\infty} e^{-\frac{v^2}{2\sigma_v^2}} \left(1 + \left(\frac{\beta_* \psi_r - \alpha_* v + \gamma_*}{B_v}\right)^2\right)^{-1} dv}. \quad (4.28)$$

Le dénominateur du membre de droite de cette équation doit être calculé numériquement.<sup>6</sup> On peut montrer que, dans la limite  $\sigma_v \rightarrow +\infty$ ,  $P(v_r | \psi_r, \beta_*, \gamma_*)$  redevient une forme purement lorentzienne.

*Fusion des distributions de probabilité* – Nous pouvons maintenant établir la probabilité conditionnelle “élémentaire” pour un objet  $i$  d'obtenir une distance mesurée  $\mu_i$  étant donné que le déplacement

<sup>6</sup>Cette fonction est connue sous le nom de profil de Voigt.

reconstruit est  $\psi_{r,i}$ , son redshift est  $z_i$ , l'erreur sur la mesure de la largeur de raie est  $\sigma_{0,i}$  et que les paramètres du modèle sont  $p'$ , dans les notations de cette section :

$$P(\mu_i|\psi_{r,i}(p_0), z_i, \sigma_{0,i}, D', p') = \int \int_{v, \mu_r} P(\mu_i|\mu_r, \sigma_{0,i}, D', p') P(\mu_r|v, z_i, p') P(v|\psi_{r,i}(p_0), p') d\mu_r dv$$

$$\propto \frac{eH}{\sigma_{0,i}} \int_{\mu_r} 10^{\mu_r/5} e^{-\frac{e^2}{2\sigma_{0,i}^2}(\mu_r - \mu_i + D - D')^2} \times P\left(v = (z_i - 10 \text{ pc} \times H 10^{\mu_r/5})|\psi_{r,i}(p_0), p'\right) d\mu_r, \quad (4.29)$$

avec  $p' = (H, \beta', B_v, \sigma_v, \gamma_*)$ ,  $\psi_{r,i}(p_0)$  étant calculé en supposant les paramètres  $p_0$ . En regardant attentivement cette probabilité, on peut remarquer que changer  $D \rightarrow D' = D + \Delta$  est équivalent à changer  $H \rightarrow H' = H \exp(\Delta/5)$ . Donc l'incertitude dans la calibration de point zéro est transféré uniquement sur  $H$ , par accroissement de son incertitude, et non sur les paramètres du modèle.

Nous pouvons à présent écrire l'expression formelle complète de  $\mathcal{L}(p)$ , comme nous l'avons déjà esquissé dans l'équation (4.20). Comme il est spécifié dans la discussion, nous prenons  $P(D')$  comme un distribution gaussienne centré sur  $D$  et avec une déviation standard  $\sigma_D$ . Maintenant, nous pouvons remplacer et obtenir :

$$\mathcal{L}'_{\beta_0}(p') \propto \int_{D'=-\infty}^{\infty} e^{-\frac{(D-D')^2}{2\sigma_D^2}} \prod_i P(\mu_i|\psi_{r,i}(p_0), z_i, \sigma_{0,i}, e, p') dD, \quad (4.30)$$

avec  $i$  parcourant l'ensemble des objets du catalogue. En supposant un prior uniforme sur  $\beta$ ,  $H$  et  $\gamma_*$  et en prenant soin de la relation entre  $\mathcal{L}'_{\beta}$  et  $\mathcal{L}$  comme mentionné précédemment, le théorème de Bayes nous permet d'écrire

$$P(H, \beta, \gamma_*|\mathfrak{M}, \mathfrak{S}, \mathfrak{Z}, e, D', B_v, \sigma_v) \propto \mathcal{L}(p) = \mathcal{L}'_{\beta}(p). \quad (4.31)$$

Nous avons à présent accès à la distribution postérieure de  $(H, \beta, \gamma_*)$ .

### 4.3.2 Méthode numérique de calcul de la fonction de vraisemblance

L'expression de la vraisemblance à un point de mesure (4.29) est délicate à évaluer numériquement car il s'agit de deux intégrales avec de longues queues de distribution. Nous avons donc procédé en différentes étapes :

- nous avons d'abord tabulé la fonction de Voigt qui, malheureusement, n'est pas très connue dans les bibliothèques mathématiques standard. De plus, il s'agit ici d'une fonction de Voigt généralisée à 2 paramètres :

$$V(x, y) = \int_{-\infty}^{+\infty} du \frac{1}{1 + (xu + y)^2} e^{-u^2/2}, \quad (4.32)$$

où usuellement la forme de Voigt est définie avec  $y = 0$ . Cette intégrale est calculée par la bibliothèque GSL (Galassi et al., 2003, <http://www.gnu.org/software/gsl/>) grâce à la fonction `gsl_integration_qag`. Comme les variations de la fonction sont suffisamment lisses, la bibliothèque GSL est en mesure de détecter l'intervalle sur lequel cette fonction prend des valeurs non-négligeables. Une fois tabulée, nous interpolons bilinéairement cette fonction suivant  $x$  et  $y$  sur le domaine de définition de la table. Avec cette définition on peut exprimer la probabilité (4.28) :

$$P(v_r|\psi_r, p) = \frac{e^{-\frac{v_r^2}{2\sigma_v^2}} \left(1 + \left(\frac{\beta_*\psi_r - \alpha_*v_r + \gamma_*}{B_v}\right)^2\right)^{-1}}{\sigma_v V(-\alpha_*\sigma_v/B_v, (\beta_*\psi_r + \gamma_*)/B_v)} \quad (4.33)$$

- Il reste à effectuer la deuxième intégrale. Malheureusement le produit sous le signe intégral dans (4.29) varie beaucoup plus rapidement que la fonction ci-dessus. Nous allons donc changer les bornes d'intégration de  $[-\infty; +\infty]$  à  $[\mu_i - D + D' - 3\sigma_{0,i}/e; \mu_i - D + D' + 3\sigma_{0,i}/e]$  afin d'intégrer

99% de l'aire de la gaussienne. Cette approximation devrait être suffisante car il y a aussi une coupure due à la probabilité (4.28). Cette intégration est de nouveau effectuée grâce à la fonction `gsl_integration_qag` de la bibliothèque GSL.

Cette méthode est suffisamment rapide pour obtenir en un temps raisonnable la fonction de vraisemblance à deux paramètres  $(h, \beta)$ . Nous allons à présent tester l'ensemble de la méthode sur des données purement aléatoires.

### 4.3.3 Test de la méthode sur des données générées aléatoirement

Les deux méthodes que nous avons utilisées pour l'estimation de la pente sont connues pour être biaisées. Un traitement plus précis de ce biais reste encore à faire. Cependant, nous nous proposons ici de vérifier quel est l'ordre de grandeur de cet effet systématique de l'analyste statistique elle-même. Afin de l'étudier, nous avons produit un ensemble de “vitesses”  $v$  générées aléatoirement et les “vitesses reconstruites” correspondantes  $v_R$ . La probabilité pour un point  $(v, v_R)$  d'avoir une vitesse  $v$  est donnée par

$$P_v(v) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-v^2/(2\sigma_v^2)}, \quad (4.34)$$

avec  $\sigma_v = 300 \text{ km s}^{-1}$  typiquement. La probabilité d'avoir une vitesse reconstruite  $v_R$  est donnée par la même loi de probabilité. Numériquement, une manière simple de fabriquer des paires de nombres aléatoires distribués selon une loi gaussienne est :

- On tire deux nombres  $a$  et  $b$  distribuée uniformément dans  $(0; 1]$ .
- On forme les quantités :

$$\begin{aligned} r &= \sqrt{-2\ln a} \\ x_1 &= r \cos(2\pi b) \\ x_2 &= r \sin(2\pi b) \end{aligned}$$

- Les variables  $x_1$  et  $x_2$  sont distribués selon une loi gaussienne de variance 1. Pour obtenir d'autres variances, il suffit de les multiplier par la variance désirée.

Nous calculons à présent l'erreur  $e$  entre  $v_R$  et  $v$ , qui doit être distribuée d'après la forme lorentzienne

$$P_{\text{DE}}(e) = \frac{1}{\pi B} \frac{1}{1 + \left(\frac{e}{B}\right)^2}, \quad (4.35)$$

avec  $B = 86 \text{ km s}^{-1}$ . Cette erreur  $e$  est liée à  $v$  et  $v_R$  par

$$e = \alpha_* v - \beta_* v_R. \quad (4.36)$$

Pour le reste de cette section, nous allons prendre  $\alpha_* = \beta_* = 1/\sqrt{2}$ . La probabilité de garder un point  $(v, v_R)$  avec une erreur  $e$  est donnée, en intégrant  $P_{\text{DE}}(e')$  entre  $-e$  et  $+e$ , par

$$P_{\text{keep}}(e) = \frac{2}{\pi} \tan^{-1} \left( \frac{e}{B} \right) \quad (4.37)$$

Nous pouvons ici appliquer une méthode de réjection.

Nous avons représenté sur la figure 4.4 un diagramme de dispersion de 10,000 points générés en utilisant la procédure ci-dessus. Comme on pourra le constater, ce diagramme ressemble à une vraie dispersion de reconstruction en redshift (cf. figure 5.9) mise à part les queues anti-diagonales. Si on conduit l'analyse à  $1.5\sigma$  sur cet ensemble de points, on trouve une pente  $\beta_*/\alpha_* = 1.0 \pm 0.20$ . En terme de mesure de la densité  $\Omega_m$ , cela donnerait  $\Omega_m = 0.30 \pm 0.10$ . En estimant la pente avec une approche en maximum de vraisemblance, cela donne, avec  $\sigma_v = +\infty$ ,  $\beta_*/\alpha_* = 0.81 \pm 0.01$  (soit  $\Omega_m = 0.20 \pm 0.02$ ) et avec  $\sigma_v = 300 \text{ km s}^{-1}$ ,  $\beta_*/\alpha_* = 1.074 \pm 0.012$  (soit  $\Omega_m = 0.34 \pm 0.02$ ). En mettant  $B = 40 \text{ km s}^{-1}$ , pour à la fois les données générées et pour la fonction de vraisemblance, comme pour les reconstructions

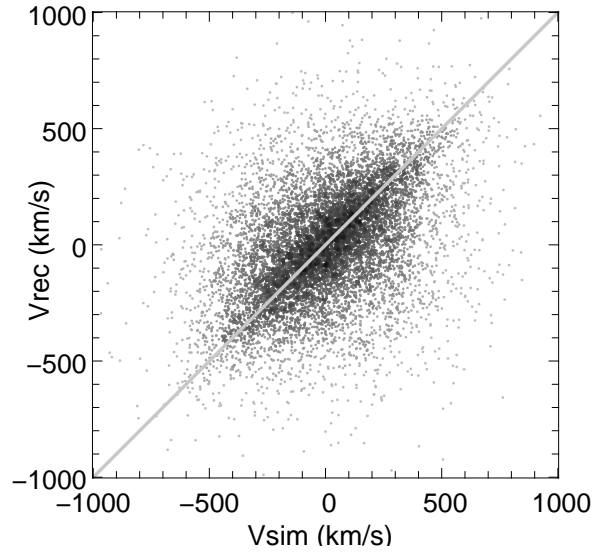


FIG. 4.4: *Biais statistique* – Diagramme de dispersion de 10 000 points générés aléatoirement suivant les lois de probabilité approchées découvertes entre les vitesses simulées et reconstruites.

en espace réel, on réduit l'erreur et cela donne  $\Omega_m = 0.31 \pm 0.02$ , ce qui met en évidence l'importance du bruit de reconstruction pour avoir une bonne estimation de  $\Omega_m$ .

Par conséquent, bien qu'on doive s'appuyer sur une analyse en fonction de vraisemblance, elle peut être fortement biaisée par la structure des erreurs de reconstruction mélangées avec une distribution non-uniforme des observables. Nous avons essayé d'avoir un modèle approché des erreurs, bien qu'il semble dépendre assez de la valeur de  $\sigma_v$ . Dans tous les cas où c'est possible on doit bien sûr faire une vérification croisée des résultats de la fonction de vraisemblance avec d'une inspection visuelle du diagramme de dispersion.

#### 4.3.4 Analyse bayésienne : introduction de la corrélation

Dans la section 4.3.1, on a supposé que les vitesses de tous les traceurs  $i$  étaient statistiquement indépendantes. C'est une hypothèse discutable. Est-il possible de faire mieux ? Si on supprime cette hypothèse nous devons réécrire la probabilité  $P(\mathcal{V}|\mathfrak{P}(p_0))$  mais on ne peut plus rien factoriser. En effet, cette probabilité s'écrit :

$$P(\mathcal{V}|\mathfrak{P}(p_0)) = \frac{P(\mathcal{V})P(\mathcal{E}(\mathcal{V}, \mathfrak{P}(p_0)))P(\mathfrak{P}(p_0))}{\int_{\mathcal{V}} P(\mathcal{V})P(\mathcal{E}(\mathcal{V}, \mathfrak{P}(p_0)))P(\mathfrak{P}(p_0))} \quad (4.38)$$

$$= \frac{P(\mathcal{V}) \prod_{i=1}^N P(e_{\psi,v,i})}{\int_{\mathcal{V}} P(\mathcal{V}) \prod_{i=1}^N P(e_{\psi,v,i})}, \quad (4.39)$$

avec  $P(\mathcal{V})$  la probabilité que  $\mathcal{V}$  soit une bonne réalisation de champ gaussien déterminé par un spectre de puissance  $P(k)$  donné,  $P(\mathfrak{P}(p_0))$  est cette même probabilité pour l'ensemble  $\mathfrak{P}(p_0)$ . On a de plus introduit la probabilité  $P(\mathcal{E})$ , qu'il y ait une erreur entre le champ de déplacement et le champ de vitesse. Ces erreurs peuvent être supposées indépendantes, ce qui nous amène à la factorisation ci-dessous :

$$P(\mathcal{E}(\mathcal{V}, \mathfrak{P}(p_0))) = \prod_{i=1}^N P(e_{\psi,v,i}), \quad (4.40)$$



avec  $e_{\psi,v,i}$  comme définit à la section 4.3.1. La théorie des champs gaussien (Bardeen et al., 1986) nous permet d'écrire la probabilité  $P(\mathcal{V})$  :

$$P(\mathcal{V}) = \frac{\sqrt{\det(C)}}{(2\pi)^{N/2}} \exp \left( \frac{1}{2} \sum_{i,j=1}^N c_{i,j} v_{r,i} v_{r,j} \right), \quad (4.41)$$

avec  $c_{i,j}$  l'élément à la ligne  $i$  et la colonne de  $j$  de la matrice  $C$ . Les éléments  $c_{i,j}$  se calculent à l'aide des fonctions de corrélation du champ de vitesse (annexe C) :

$$c_{i,j} = \langle v_{r,i} v_{r,j} \rangle = \sum_{l,m} \hat{x}_i^l \hat{x}_j^m \xi_{v,l,m}(\mathbf{x}_i, \mathbf{x}_j), \quad (4.42)$$

avec  $\hat{x}_b^a$  la composante  $a$  du vecteur position du traceur  $b$ ,  $\xi_{v,l,m}(\mathbf{y}, \mathbf{z})$  la composante  $(l, m)$  du tenseur des corrélations du champ de vitesse pris entre les points de position  $\mathbf{y}$  et  $\mathbf{z}$ . On peut exprimer cette fonction à l'aide du spectre de puissance des fluctuations de densité  $P(k)$  (cf. section C.2) :

$$\xi_{v,l,m}(\mathbf{y}, \mathbf{z}) = \frac{(\beta H)^2}{(2\pi)^3} \int_{\mathbf{k}} d^3\mathbf{k} \frac{k_l k_m}{k^4} P(k) e^{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{y})}, \quad (4.43)$$

$$= \frac{(\beta H)^2}{(2\pi)^3} \int_{k=0}^{+\infty} dk \int_{\Omega} d\Omega(\hat{k}) \hat{k}_l \hat{k}_m e^{ik\hat{k} \cdot (\mathbf{y} - \mathbf{z})}, \quad (4.44)$$

avec  $\hat{k} = \mathbf{k}/\|\mathbf{k}\|$ ,  $\Omega$  l'angle solide correspondant à la direction de  $\hat{k}$ . Comme cette expression est invariante par translation de l'espace, on pose alors :

$$\xi_{v,l,m}(\mathbf{y}, \mathbf{z}) = \xi_{l,m}^*(\mathbf{z} - \mathbf{y}), \quad (4.45)$$

et on va chercher à calculer  $\xi_{l,m}^*$ . Pour calculer cette expression, on peut poser l'*ansatz* :

$$\xi_{l,m}^*(\mathbf{r}) = \delta_{l,m} A(\|\mathbf{r}\|) + \hat{r}_l \hat{r}_m B(\|\mathbf{r}\|). \quad (4.46)$$

On remarque alors que :

$$\sum_l \xi_{l,l}^*(\mathbf{r}) = 3A + B = \frac{(\beta H)^2}{(2\pi)^3} \int_{k=0}^{+\infty} dk P(k) \int_{\Omega} d\Omega(\hat{k}) e^{ik\hat{k} \cdot (\mathbf{r})} \quad (4.47)$$

$$= \frac{(\beta H)^2}{2\pi^2} \int_{k=0}^{+\infty} dk P(k) j_0(k\|\mathbf{r}\|), \quad (4.48)$$

$$\sum_{l,m} \hat{r}_l \hat{r}_m \xi_{l,m}^*(\mathbf{r}) = A + B = \frac{(\beta H)^2}{(2\pi)^3} \int_{k=0}^{+\infty} dk P(k) \int_{\Omega} d\Omega(\hat{k}) (\hat{k} \cdot \mathbf{r})^2 e^{ik\hat{k} \cdot (\mathbf{r})} \quad (4.49)$$

$$= \frac{(\beta H)^2}{2\pi^2} \int_{k=0}^{+\infty} dk P(k) \left( \frac{j_0 - 2j_2}{3} \right) (k\|\mathbf{r}\|), \quad (4.50)$$

avec  $j_0$  et  $j_2$  les fonctions de Bessel défini dans l'annexe A.4. On trouve donc :

$$A(r) = \frac{(\beta H)^2}{6\pi^2} \int_{k=0}^{+\infty} P(k) (j_0 + j_2)(kr) \quad (4.51)$$

$$B(r) = -\frac{(\beta H)^2}{2\pi^2} \int_{k=0}^{+\infty} P(k) j_2(kr) \quad (4.52)$$

Au final, si on rassemble tous les termes, le coefficient  $c_{i,j}$  vaut

$$c_{i,j} = (\hat{x}_i \cdot \hat{x}_j) A(\|\mathbf{x}_i - \mathbf{x}_j\|) + \left( \hat{x}_i \cdot \widehat{(\mathbf{x}_i - \mathbf{x}_j)} \right) \left( \hat{x}_j \cdot \widehat{(\mathbf{x}_i - \mathbf{x}_j)} \right) B(\|\mathbf{x}_i - \mathbf{x}_j\|) \quad (4.53)$$

On voit donc que cette fois on ne peut plus faire  $N$  fois une intégrale pour calculer  $P(\mathcal{V}|\mathfrak{P})$  mais une  $N$ -intégrale sur  $[-\infty; +\infty]$  car  $P(\mathcal{V})$  ne peut plus être factorisé. Par ailleurs, on ne peut pas non plus réorganiser les termes avec ceux de  $P(e_{\psi,v})$  car cette fonction n'est pas strictement gaussienne. Faute de temps, il ne nous a pas été possible de pousser les investigations plus avant en suivant cette approche qui peut néanmoins s'avérer prometteuse afin de limiter notre sensibilité au bruit de reconstruction et de mesure. On peut citer par exemple Zaroubi et al. (1997) qui a tenté une approche similaire sur le champ de vitesse. Plusieurs pistes pourront être tentées dans le futur :

1. Revenir à une formulation plus simple de  $P(e_{\psi,v})$  et ne faire intervenir que des gaussiennes dans l'expression.
2. Vérifier si les  $c_{i,j}$  ne peuvent pas être mis en paquets semi-indépendants. Cela permettrait de réduire la dimensionnalité de l'intégrale. L'idéal serait de ne pas avoir une intégrale de dimension plus grande que 3 ou 4 pour que cela reste gérable numériquement.

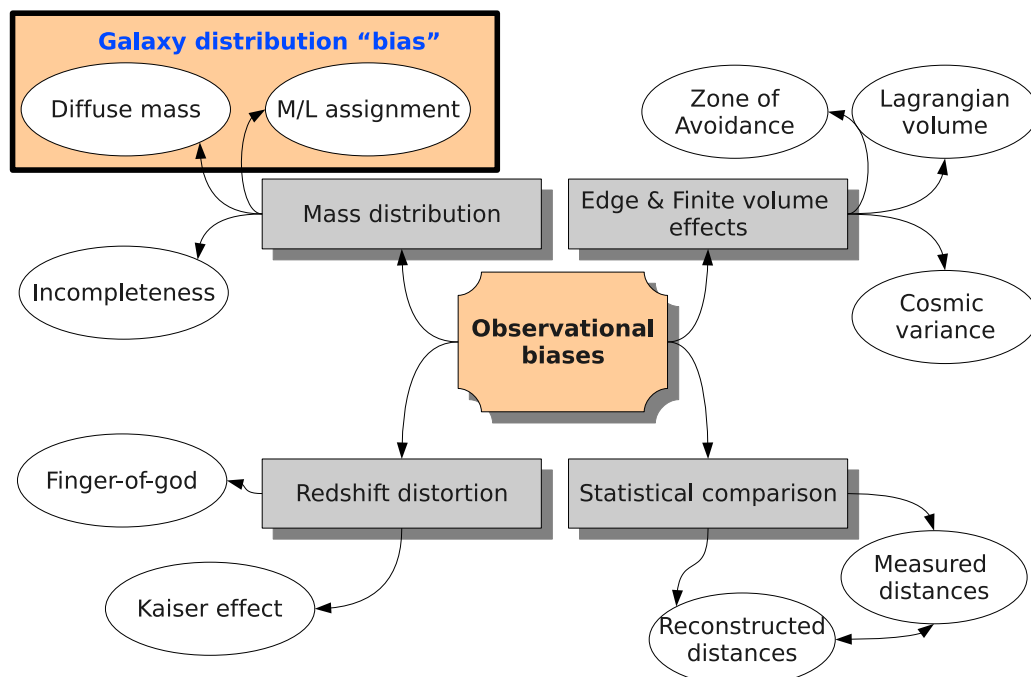


# 5 Etude des biais observationnels avec des catalogues virtuels de galaxies

## Sommaire

<b>5.1</b>	<b>Catalogues virtuels de base pour l'étude</b>	<b>68</b>
<b>5.2</b>	<b>La relation masse-luminosité</b>	<b>69</b>
5.2.1	Absence de traceurs de masse	70
5.2.2	Rapport masse-luminosité de groupes de galaxies	72
5.2.3	Limitation en magnitude	76
<b>5.3</b>	<b>Distortion de décalage vers le rouge (redshift)</b>	<b>79</b>
<b>5.4</b>	<b>Effets dues à la géométrie du catalogue</b>	<b>86</b>
5.4.1	Le plan galactique de la voie lactée	86
5.4.2	Le domaine lagrangien	87
5.4.3	La variance cosmique	93
<b>5.5</b>	<b>Erreurs sur les mesures de distance</b>	<b>98</b>
5.5.1	A-t-on réellement besoin d'une analyse de vraisemblance?	98
5.5.2	Résultat	98
<b>5.6</b>	<b>Conclusion sur les systématiques</b>	<b>101</b>

DANS les reconstructions du chapitre 3, nous nous sommes bornés à considérer des catalogues de halos “parfaits”. En effet, on avait un échantillonnage homogène, les positions et les masses. En effet, on avait typiquement un échantillonnage homogène, des positions et des masses des traceurs connues, les effets de bord étaient absents. Malheureusement, un vrai catalogue de galaxies, du type de ceux que nous allons présenter au chapitre 6, ne sont malheureusement absolument pas parfaits. Il contient de plus de nombreux biais observationnels. Il va donc nous falloir tester la robustesse de la reconstruction MAK à l’aide de catalogues virtuels plus réalistes. Nous nous proposons d’étudier les problèmes qui sont schématisés dans le graphe ci-dessous :



Cette liste n'est pas exhaustive mais permet déjà d'identifier un certain nombre de problèmes importants :

- Tout d'abord celui qui va nous poser, en définitive, le plus de problème : l'affectation de la masse sur des traceurs lumineux (section 5.2). Généralement ce problème est aussi connu de manière plus général comme le biais de la distribution de galaxies d'un catalogue par rapport à la distribution de matière noire. Le biais  $b$  est défini à partir de la relation suivante

$$\delta_{\text{gal}}(\mathbf{x}) \simeq \frac{1}{b} \delta_{\text{m}}(\mathbf{x}) = \mathcal{F}(\delta_{\text{lum}}(\mathbf{x})) , \quad (5.1)$$

avec  $\delta_{\text{gal}}$  les fluctuations de densité numérique de galaxies,  $\delta_{\text{m}}$  les fluctuations de densité de matière (noire+baryonique),  $\delta_{\text{lum}}$  les fluctuations de densité de luminosité. Dans la littérature, on s'intéresse principalement à étudier le coefficient qui relie les deux quantités de gauche. Nous nous intéresserons ici plutôt à l'égalité de droite et sur la fonctionnelle  $\mathcal{F}$ , qui peut dépendre de la distance à l'observateur.

- Le deuxième aspect important de ce chapitre est l'étude de l'impact des distortions redshift de “grande” échelle (section 5.3).
- Le troisième aspect est l'étude de l'impact des effets de bord au sens large. Nous en avons isolé trois : un effet dû à l'obscurcissement d'une région du ciel par le plan galactique de la Voie Lactée et la caractéristique fini des volumes de catalogues de galaxies qui se traduit par une méconnaissance du volume lagrangien nécessaire pour la reconstruction et des effets de variances cosmiques (section 5.4)
- Enfin, nous discuterons des problèmes de comparaisons des vitesses observées aux vitesses reconstruites (section 5.5).

## 5.1 Catalogues virtuels de base pour l'étude

Bien que beaucoup de recettes seront employées par la suite pour étudier les différents biais observationnels, nous partirons toujours des trois<sup>1</sup> mêmes catalogues de halos “principaux”. Le premier catalogue tâche de reproduire, jusqu'à un certain point, les principales caractéristiques de notre Univers local, en particulier la présence d'un gros amas à environ  $40h^{-1}$  Mpc et d'un super-amas à environ  $70h^{-1}$  Mpc. Les deuxième et le troisième catalogues n'ont pas ces caractéristiques mais représentent des réalisations localement sur- et sous- dense afin d'analyser le problème de la variance cosmique.

A partir de *FullMock*, nous avons donc extrait trois coupes sphériques de rayon  $40 h^{-1}\text{Mpc}$  (par la suite nommé *4k-mockX*), où les comparaisons vitesse-vitesse sont effectuées, et leur homologue deux fois plus profond (par la suite nommé *8k-mockX*) qui seront utilisés pour donner de meilleures contraintes (section 5.4.2) sur la reconstruction dans le volume de  $40 h^{-1}\text{Mpc}$ . Chacun de ces catalogues est centré dans un endroit différent de la simulation d'origine de telle manière que :

- *4k-mock6* est légèrement sur-dense, avec une densité moyenne de la matière  $\Omega_{\text{eff}} = 0.35$ , et contient 495 halos. Il a été conçu de manière à ce que de grandes zones vides et des zones à forte concentration de matière (amas et super-amas) soient présentes au voisinage de sa frontière, de manière similaire à ce qu'on trouve dans les catalogues redshift représentant notre Univers local, tel que le UZC (Falco et al., 1999), le NBG-3k (Shaya et al., 1995; Tully et al., 2008) et le NBG-8k. Ce catalogue et son équivalent plus profond, *8k-mock6*, sont particulièrement adaptés à l'étude des effets de bord dans le NBG-3k (qui s'arrête aux amas de l'Hydra et de Centaurus) et le NBG-8k (qui s'arrête au Great Wall), respectivement.
- *4k-mock7* est très sur-dense, avec  $\Omega_{\text{eff}} = 0.50$ , et contient 656 halos. Une faible quantité de masse entre et sort de ce volume entre l'instant des conditions initiales et la fin de la simulation : ce volume se comporte comme un Univers isolé, avec peu d'effets de marée induits par la distribution de masse extérieure.

<sup>1</sup>Le coût d'une reconstruction est absolument rédhibitoire pour une étude exhaustive de toutes les réalisations possibles d'Univers. Nous nous limitons donc aux quelques catalogues mentionnés ici. Typiquement sur un quadri-processeur AMD Opteron il faut de l'ordre de un mois pour résoudre le problème à  $128^3$  particules.

- *4k-mock12* est sous-dense, avec  $\Omega_{\text{eff}} = 0.19$ , et contient 213 halos. Il a aussi peu de fluctuations de densité sur sa frontière.

## 5.2 La relation masse-luminosité

La plupart des méthodes de reconstruction, y compris celle que nous utilisons ici, déduisent la distribution de matière totale à partir de la distribution de luminosité visible tracée par les galaxies. L'hypothèse fondamentale derrière cette approche est que la relation entre ces deux quantités est très déterministe. Autrement dit, on affecte à chaque galaxie de luminosité  $L$  une concentration de matière noire (un halo) de masse  $M = f(L)$ . Cependant, cette procédure comporte plusieurs points durs dont les deux plus importants sont les suivants :

- *le rapport masse-sur-luminosité* – Le choix de la fonction  $f(L)$  influence considérablement les résultats. On s'attend à introduire un biais significatif sur la mesure de  $\beta$  si ce choix est inadéquat. Pour le moment, la fonction  $f(L)$  est grossièrement déterminée (Tully, 2005; Marinoni & Hudson, 2002) à partir des observations. Une façon de déterminer cette fonction est de s'appuyer sur les modèles semi-analytiques de formation des galaxies, mais ceci représente un a-priori très important sur les mesures. De plus,  $f(L)$  ne reste qu'une relation moyenne autour de laquelle il peut y avoir une dispersion importante. Cette dispersion peut aussi bien introduire des biais significatifs.
- *Absence de traceurs / Limitation en magnitude* – Même si la fonction  $f(L)$  est connue parfaitement, les galaxies plus faibles sont néanmoins absentes dans les catalogues à cause de limitations dans les instruments observationnels. Par exemple, dans les catalogues limités en magnitude, la densité numérique de galaxies détectées décroît avec la distance à l'observateur. Ces traceurs manquants ont une position inconnue et correspondent à une partie de la distribution de matière noire qui reste totalement indéfinie. On doit tenir compte de cette masse manquante d'une manière ou d'une autre.

Dans ce qui suit, nous allons d'abord résoudre le deuxième point d'une manière très simple en supposant que la fonction  $f(L)$  est connue (c'est-à-dire en fait la masse des halos de matière noire eux-mêmes) mais qu'il y a une coupure fixe à petite masse. Le problème consiste à déterminer la part inconnue de la distribution de matière noire (c'est-à-dire les particules qui ne sont affectées à aucun halo). Clairement, elle est corrélée aux traceurs restant dans le catalogue mais elle est moins concentrée. Il y a deux manières extrêmes de localiser cette masse manquante :

- (a) l'associer à des traceurs existants comme c'est généralement réalisé dans les analyses sur des observations réelles.
- (b) l'associer à un champ d'arrière plan uniformément réparti.

Bien sur, la vraie solution se situe quelque part entre (a) et (b) comme cela le sera montré dans la section 5.2.1.

Par la suite, nous nous tournerons, dans la section 5.2.2, vers le problème de choisir un  $f(L)$ . Dans ce travail, nous préférons être aussi détachés que possible d'a priori fort, donc nous n'utilisons délibérément pas les résultats des modèles semi-analytiques de formation de galaxies. A la place, nous utilisons la détermination de  $f(L)$  à partir des données observationnelles mais, malheureusement, avec une énorme incertitude sur ces mesures. Le point important est de quantifier ici, quoiqu'assez empiriquement, l'effets de ces incertitudes, aléatoires et systématiques, sur la mesure de  $\beta$ . En effet, on est à la fois confronté à une possibilité d'une mauvaise approximation de  $f(L)$  et très probablement à une énorme dispersion autour de cette relation moyenne.

Dans les catalogues de galaxie suffisamment profonds, on s'attend à ce que les effets dûs aux manques de traceurs soient négligeables au voisinage de l'observateur et, en général, à augmenter avec la distance à l'observateur. Avec une bonne pondération des données, on peut minimiser le biais apporté par cette procédure utilisée pour déduire la distribution de la masse manquante à grande distance de l'observateur. Dans la section 5.2.3, nous illustrerons ce point en considérant le cas des catalogues limités en magnitude où toute la masse manquante est associée aux traceurs observés [méthode (a) ci-dessus].

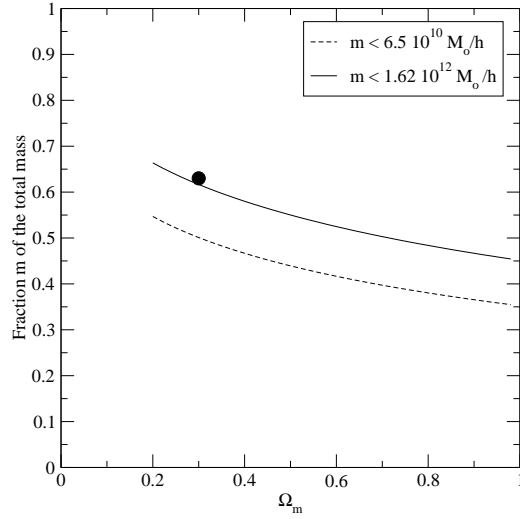


FIG. 5.1: *Masse diffuse* – Sur ce graphique, nous représentons la fraction de la masse non-effondrée pour deux résolutions de masse différentes pour une cosmologie standard type  $\Lambda$ CDM ( $h = 0.65$ ,  $\sigma_8 = 1.0$ ). Nous avons utilisé une forme de spectre de puissance comme il a été proposé par Bardeen et al. (1986). La courbure de l’Univers reste plate alors que  $\Omega_m$  varie. La fraction est représentée pour deux résolution en masse :  $2.5 \times 10^{12} M_\odot$  (ce qui correspond à la limite basse de la masse des halos dans notre simulation) et  $6.5 \times 10^{10} M_\odot$  ( $\simeq 10^9 L_{\odot,B}$ ). La fraction de masse non-effondrée dans *FullMock* est donnée par le cercle rempli noir. La fraction de masse en-dessous chacune de ces deux limites est encore considérable.

### 5.2.1 Absence de traceurs de masse

La figure 5.1 illustre la prédiction, obtenue en utilisant le modèle de Sheth & Tormen (2002) (voir aussi figure 3.2 et annexe D), de la fraction attendue de masse sous un seuil donné comme une fonction de  $\Omega_m$ . La ligne solide qui correspond à la coupure en masse des halos dans *FullMock* est en bon accord avec la mesure obtenue dans la simulation pour  $\Omega_m = 0.30$ . Dans ce cas, 63% de la masse est resté en dehors des halos, qui représentent nos “galaxies” avec un rapport  $M/L$  connu. Les particules qui ne sont liées à aucun halos représentent la masse manquante. Sur la figure 3.9, leur position exacte à été utilisée pour effectuer la reconstruction. La seule information disponible est maintenant la distribution des “galaxies visibles”. La masse manquante doit alors être redistribuée en utilisant seulement ces éléments d’informations (c’est-à-dire la fraction approximative de masse manquante). Nous proposons deux façons extrêmes de traiter le problème :

- (I) *Toute la masse manquante est mise dans le champ de masse d’arrière plan* – Avant la reconstruction, la masse manquante est divisée en particules qui sont placées aléatoirement dans le volume de la simulation en suivant une statistique de Poisson. Dans l’exemple illustré par les panneaux de droite de la figure 5.2, nous avons choisi par simplicité des particules de la même masse que celles utilisées dans la simulation ;
- (II) *Toute la masse manquante dans les halos* – La masse manquante est attribuée aux halos existants d’une manière proportionnelle à leur masse, comme illustré dans les panneaux de gauche de la figure 5.2. Cette approche est équivalente, dans les observations réelles, à multiplier le rapport  $M/L$  pour les galaxies ou les groupes de galaxies par une constante  $\alpha > 1$ .

Bien sûr, dans le cas (I), l’effet d’écrantage dû au champ d’arrière-plan est exagéré ; par conséquent les vitesses reconstruites sont moins contrastées et  $\beta$  est sur-estimé pour compenser cet effet. Dans le cas (II), au contraire, les puits de potentiel sont plus contrastés qu’ils ne le devraient, ce qui conduit à l’effet opposé. Parvenu à ce point, il est extrêmement tentant d’essayer de trouver le compromis optimal entre (I) et (II) comme illustré dans les panneaux centraux de la figure 5.2 où 60% de la

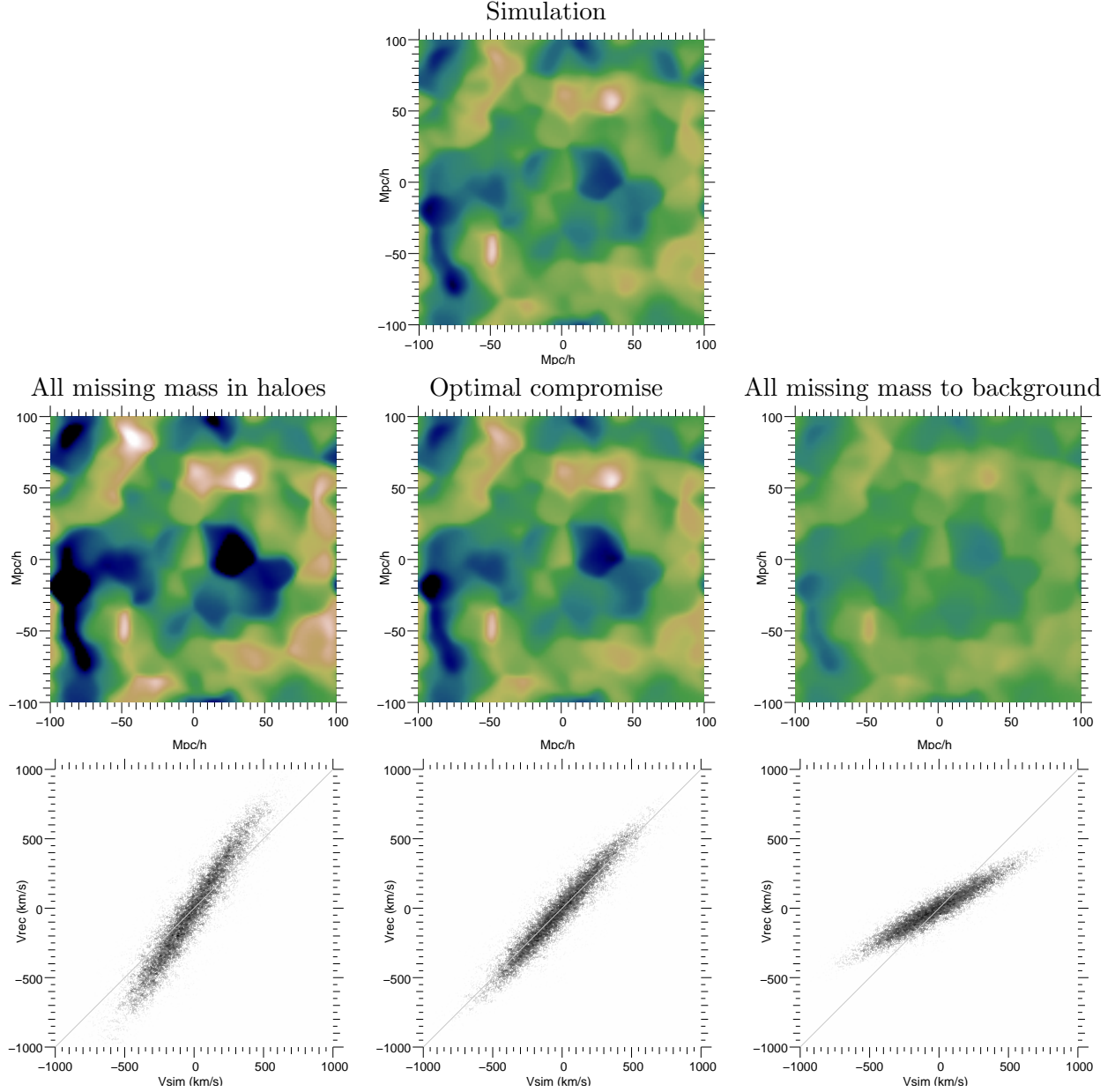


FIG. 5.2: *Correction de l'effet de masse diffuse* – Le panneau du haut donne une coupe de la composante selon la ligne de visée du champ de vitesse simulé, après lissage avec une fenêtre gaussienne de rayon  $5 h^{-1}\text{Mpc}$ . L'observateur a été placé au centre de la coupe. La deuxième ligne de panneaux représente la composante selon la ligne de visée du champ de vitesse reconstruit, lissé de la même manière, pour différentes correction de la masse diffuse. La troisième ligne de panneaux donne le distribution des vitesses reconstruites pour chacun des halos *par rapport* à celles simulées. Les panneaux de gauche montrent le résultat d'une reconstruction sur un catalogue virtuel qui contient uniquement les halos et pas le champ d'arrière-plan, mais qui à la fois conserve la masse totale du catalogue tout en réaffectant la masse manquante aux halos. Les panneaux de droites montrent le résultat d'une reconstruction basé sur un catalogue virtuel pour lequel la masse manquante diffuse est représenté par un champ d'arrière-plan composé de particules placés aléatoirement dans le catalogue. Les panneaux centraux montrent le résultat d'une reconstruction sur un catalogue virtuel qui ne contient que les halos et un champ d'arrière-plan aléatoire. La masse qui a été initialement enlevée du catalogue virtuel (les “galaxies” de champ) a été réaffectée comme suit : 60% aux halos et 40% au champ.



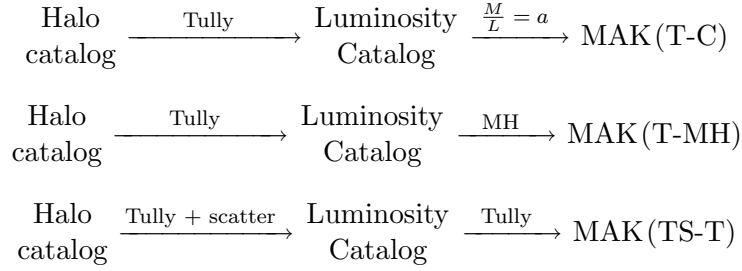


FIG. 5.3: *Rapport  $M/L$*  – Schéma de la procédure utilisé pour tester l’influence d’un choix d’affectation  $M/L$ , comme c’est expliqué dans le corps du texte.

masse manquante a été liée aux traceurs “observés” et le reste mis dans l’arrière-plan homogène. Avec ce choix particulier de la redistribution de la masse, l’accord entre les champs reconstruit et simulé est spectaculaire. Ce résultat est non-trivial étant donné la simplicité de la gestion des 63% de la masse manquante, et il l’est d’autant plus que la dispersion du panneau au milieu en bas de la figure 5.2 est du même ordre que celui en bas à gauche de la figure 3.9, où tous les traceurs contribués de la manière optimale.

Bien que ce choix d’une redistribution optimale reste a priori inconnu dans un vrai catalogue de galaxie, on peut au moins en déduire des barres d’erreur à partir des cas (I) et (II). Dans ce cadre-là, le figure 5.2 nous amène malheureusement à des contraintes faible sur  $\beta$ ,  $0.36 \lesssim \beta \lesssim 0.85$ . Cependant, dans les vrais catalogues de galaxie, comme le NBG-3k et le NBG-8k, la luminosité minimale est de l’ordre de  $10^9 L_\odot$ . Cette limite correspond à une coupure moins abrupte en masse,  $M_{\text{cut}} \sim 10^{11} M_\odot$ , que dans la figure 5.2, où  $M_{\text{cut}} = 2.5 \times 10^{12} M_\odot$ . Par conséquent, nous nous attendons à ce que la masse manquante soit un problème de moindre ampleur dans le traitement des résultats d’observations réelles comme illustré par la courbe en tiret de la figure 5.1. De plus, un usage approprié des catalogues virtuels peut aider à calibrer la redistribution de masse, comme cela a été fait pour les panneaux de milieu de la figure 5.2.

### 5.2.2 Rapport masse-luminosité de groupes de galaxies

Pour tester quel est l’impact sur le résultat du choix de l’affectation de la masse aux galaxies ou aux groupes de galaxies, nous allons considérer les trois cas suivants, résumés sur la figure 5.3 :

1. *Cas T-C* : un catalogue de galaxies est extrait de *FullMock* en associant une luminosité  $L(M)$  à chaque halo de matière noire ou à une particule de l’arrière-plan en utilisant le dernier meilleur ajustement de Brent Tully sur la relation masse-luminosité des groupes de galaxies (Tully, 2005, voir Figure. 5.4)

$$\frac{L_B}{L_\odot} = 2700 \left( \frac{M}{M_\odot} \right)^{0.59} e^{-6 \times 10^{11} M_\odot / M}, \quad (5.2)$$

qui donne la luminosité en bande B pour des groupes dans l’intervalle de masse  $10^{11} M_\odot - 10^{15} M_\odot$ . Une nouvelle masse est alors attribuée à chaque traceur en supposant

$$M/L = \text{constante}, \quad (5.3)$$

comme il est souvent fait dans la littérature scientifique. La reconstruction MAK est effectuée sur un rééchantillonnage de la nouvelle distribution de masse.

2. *Cas T-MH* : il s’agit d’un cas moins extrême que supposer  $M/L = \text{constante}$ . Il consiste à séparer les traceurs en trois grandes classes : les galaxies faibles, les galaxies lumineuses et les groupes/amas de galaxies, comme cela a été fait par Marinoni & Hudson (2002), par la suite

noté MH. Pour faire cela, ils ont utilisé la fonction de luminosité de Schechter et la fonction de masse de Press-Schechter, ce qui donne le résultat suivant :

$$\begin{aligned}
 M/L &= 1.15 \cdot 10^7 \left( \frac{L}{L_\odot} \right)^{-0.5} h \frac{M_\odot}{L_\odot} & L < 4 \cdot 10^{10} L_\odot \\
 M/L &= 128 h M_\odot & 4 \cdot 10^{10} L_\odot < L < 4 \cdot 10^{11} L_\odot \\
 M/L &= 3.6 \cdot 10^{-4} \left( \frac{L}{L_\odot} \right)^{0.5} h \frac{M_\odot}{L_\odot} & L > 4 \cdot 10^{11} L_\odot
 \end{aligned} \tag{5.4}$$

comme il est montré dans le panneau du haut de la figure 5.4. Dans ce cadre-là, nous avons généré le même catalogue que dans le cas  $T-C$  mais dont la distribution de masse a été obtenue en supposant la fonction  $M/L$  donné par l'équation (5.4).

3. *Cas TS-T* : en supposant qu'il est possible d'avoir un estimateur non-biaisé de la fonction  $M/L$ , il peut tout de même y avoir une dispersion importante par rapport à cette valeur moyenne, ce qui pourrait être la source d'une augmentation des erreurs et induire un biais. Nous testons ceci en multipliant la masse de chaque halo de *FullMock* par un nombre aléatoire  $x$  tel que  $\log_{10} x$  est uniformément distribué dans  $[-1, 1]$ , avant la reconstruction MAK, qui est effectuée sur un ré-échantillonnage du catalogue de halo en suivant la procédure expliquée à la section 3.3.2. Il faut remarquer que la masse des particules d'arrière-plan reste inchangée lors de l'application de cette procédure, ce qui correspond à ne pas toucher la distribution de 63% de la matière dans le cadre de cette nouvelle dispersion. Cependant, appliquer une dispersion uniquement aux halos de faible masse seulement introduit peu de bruit additionnel ce qui ne devrait pas avoir de conséquences importantes sur la précision de la reconstruction laquelle offre une plus grande sensibilité aux puits de potentiel plus profonds.

Il faut se rendre compte, qu'en fait, aucune de ces transformations de la masse d'un objet de *FullMock* à travers une opération  $\tilde{M} \rightarrow L \rightarrow M$ , ne correspond à l'identité. On obtient en fait un nouvel ensemble de masse attaché à chaque traceur qui est différent de l'ensemble d'origine. De plus, la distribution de masse en sortie  $P_{\text{mass,out}}(M)$  peut être fondamentalement différente de celle en entrée  $P_{\text{mass,in}}(M)$ . En effet, calculer  $P_{\text{mass,out}}(M)$  est équivalent à effectuer une déformation globale de  $P_{\text{mass,in}}(M)$ . Le calcul donne :

$$P_{\text{mass,out}}(M) dM = P_{\text{mass,in}}(\tilde{M}(M)) d\tilde{M} = P_{\text{mass,in}}(\tilde{M}(M)) \frac{d\tilde{M}}{dM} dM, \tag{5.5}$$

avec  $\tilde{M}(M)$  donnée par le diagramme de la figure 5.3. Par conséquent, les propriétés statistiques du champ de densité de masse correspondant peuvent en être affectées.

D'une manière plus technique, la conservation de la masse totale est une contrainte utilisée lors de l'application de la procédure de construction de tous les catalogues présentés ci-dessus. Il faut remarquer que la masse totale ne dépend que de  $\Omega_m h^2$ , mais cette normalisation n'affecte pas les déplacements de MAK, qui ne sont sensibles qu'aux contrastes de densité. Les paramètres  $\Omega_m$  et  $h$  n'interviennent en fait que, respectivement, dans la comparaison vitesse-vitesse et lors de la conversion des distances en vitesse (section 5.5).

Comme on s'y attend, une erreur aléatoire sur la détermination de la masse n'introduit pas de biais, elle ne fait qu'augmenter la dispersion dans les mesures comme cela est illustré par le panneau en bas à gauche de la figure 5.5. Un problème plus important est la connaissance globale de la relation  $M/L$ . En effet, il semble que la pente (logarithmique) de cette relation influence grandement le résultat, comme cela est illustré par les panneaux du milieu et de droite de la figure 5.5. Si les galaxies suivent la formule de Tully (5.2), il est clairement faux de supposer  $M/L$  constant. Il faut même ajouter que l'ajustement de MH introduit un biais significatif, même si l'écart à l'équation (5.2) est bien en dessous des erreurs observationnelles. Il convient de noter que ce biais peut être transformé en avantage si on ne souhaite pas mesurer  $\Omega_m$  mais la relation  $M/L$ . En effet, l'expérience WMAP (Bennett et al., 2003; Spergel et al., 2006) couplée à une analyse du spectre de puissance du champ de densité des galaxies à grande échelle (Tegmark et al., 2006) donne à présent de bonnes contraintes sur le vrai  $\Omega_m$ . Notre méthode,

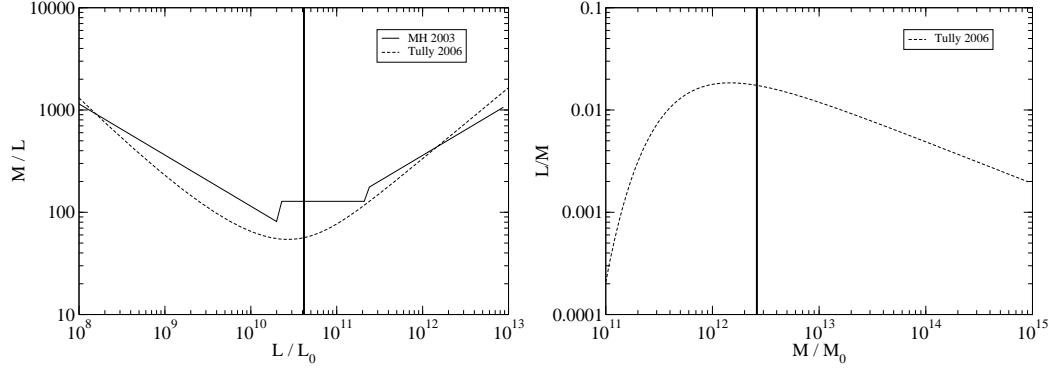


FIG. 5.4: *Fonction  $M/L$*  – Les deux graphiques ci-dessus donnent les fonctions directes ( $M/L = f(L)$ ) et inverses ( $L/M = f(M)$ ) du rapport masse-sur-luminosité pour à la fois les ajustement de Tully (2005) et Marinoni & Hudson (2002). Le panneau de gauche donne le rapport  $M/L$  comme une fonction de la luminosité  $L$ , le panneau de droite donne  $L/M$  comme une fonction de la masse  $M$ .

TAB. 5.1: *Effet de biais de  $M/L$*  – Cette table donne les résultats obtenus avec différents outils statistiques. Nous avons aussi mesuré  $\Omega_m$  avec six différentes méthodes : le label  $s$  signifie que nous avons utilisé la pente estimé en utilisant tous les objets, le label  $\mathfrak{L}$  est utilisé quant  $\Omega_m$  a été déterminé en utilisant l’analyse en fonction de vraisemblance, et le label  $1.5\sigma$  est utilisé quand la pente a été estimée en utilisant seulement les objets à l’intérieur de l’isocontour à  $1.5\sigma$  de la PDF entre les vitesses reconstruites et les vitesses simulées (méthode décrite dans la section 4.2.1).

Transf.	Velocity			$\Omega_m (s)$	$\Omega_M (\mathfrak{L}_{\min})$	$\Omega_M (\mathfrak{L}_{\max})$	$\Omega_M (1.5\sigma, s_{\text{med}})$	$\Omega_M (1.5\sigma, s_{\min})$	$\Omega_M (1.5\sigma, s_{\max})$
	$s$	$r$	$\sigma$						
None	0.88	0.89	0.58	0.38	0.30	0.31	0.30	0.28	0.31
TS-T	0.90	0.78	0.64	0.36	0.26	0.30	0.28	0.24	0.33
T-MH	0.80	0.80	0.60	0.45	0.33	0.38	0.36	0.32	0.40
T-C	0.71	0.78	0.63	0.55	0.40	0.48	0.44	0.37	0.51

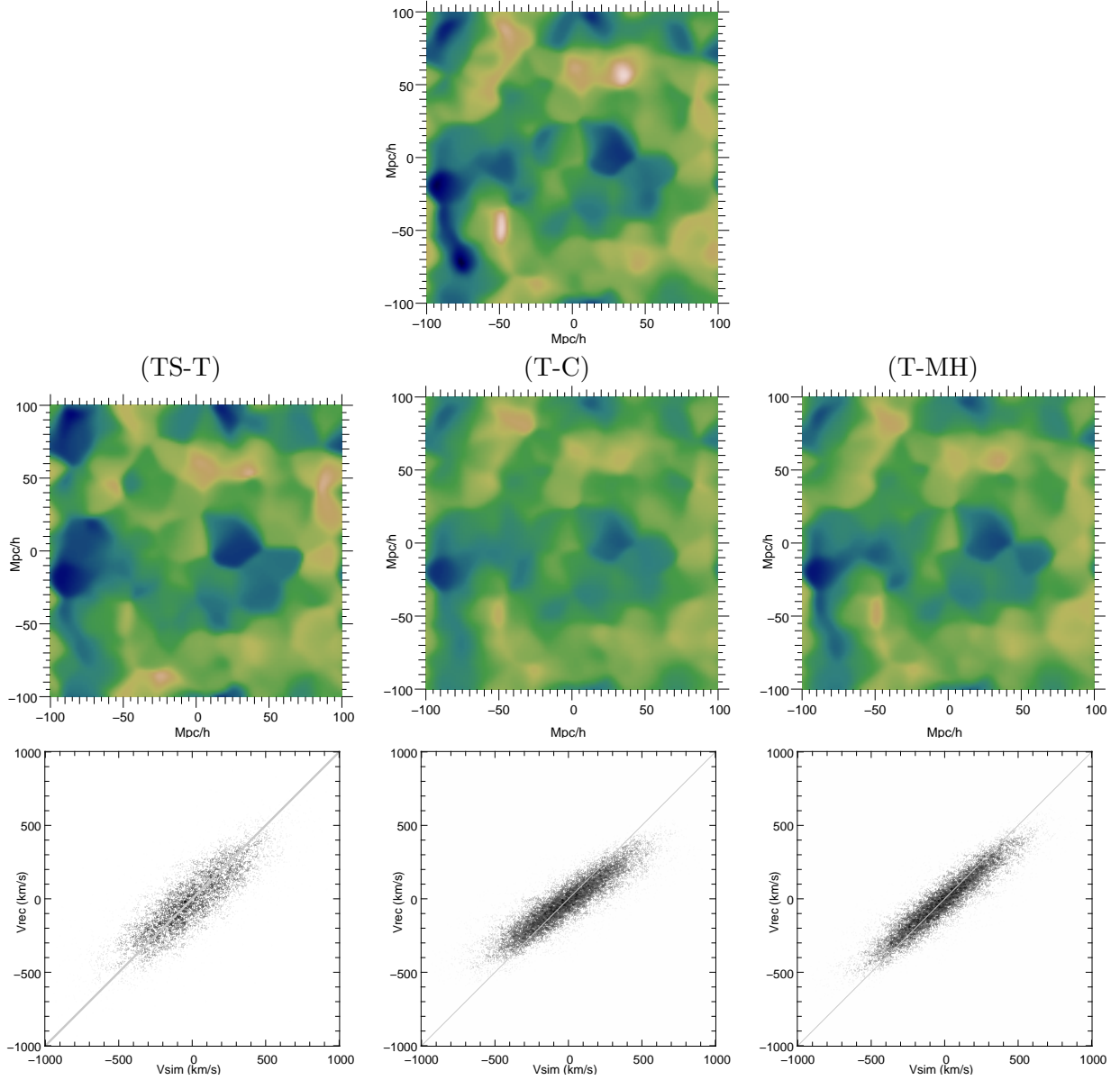


FIG. 5.5: *Le biais de  $M/L$*  – Les panneaux du haut donnent la composante attendue selon la ligne de visée  $v_r$  du champ de vitesse, lissé avec un filtre gaussien de  $5 h^{-1}\text{Mpc}$  de rayon, tel que donné par la simulation dans une tranche fine de la simulation contenant l’observateur ; Les panneaux du milieu donnent le champ  $v_r$  reconstruit, avec le même lissage, après avoir appliqué les transformations spécifiées dans la figure 5.3 à *FullMock*. Les panneaux du bas donnent la dispersion entre les vitesses reconstruites et simulées pour chacune de ces transformations.

au contraire, est capable de mesurer la différence entre le  $\beta$  mesuré et le taux de croissance attendu  $\beta_{\text{expected}} = \Omega_m^{5/9}$  (c'est-à-dire le biais). Cette mesure peut donner une idée de l'erreur commise sur la relation  $M/L$  utilisée avant la reconstruction et mettre des contraintes plus fortes sur des fonctions  $M/L$  potentiellement acceptables. Par conséquent, notre méthode est capable de mesurer la *façon* dont la matière est distribuée dans l'Univers une fois que l'on se donne sa densité moyenne  $\Omega_m$ . D'un autre côté, si le biais ci-dessus est bien compris, cette méthode permet de réduire la dégénérescence dans la détermination des paramètres cosmologiques. En effet, la distribution de probabilité postérieure donne des contraintes orthogonales (voir par exemple les résultats dans Mohayaee & Tully, 2005) à ceux obtenu dans l'expérience WMAP et la statistiques des galaxies du SDSS.

### 5.2.3 Limitation en magnitude

Un échantillon limité en magnitude introduit un nouveau type de problème : la limitation en flux fait décroître la résolution en masse sur les bords externes du catalogue contrairement au cas homogène étudié dans la section 5.2.1. Habituellement, l'incomplétude d'un catalogue est corrigée en amplifiant uniformément les luminosités des galaxies à une certaine distance de l'observateur (Branchini et al., 2002), avant la conversion des luminosités en masses. C'est une approche convenable si  $M/L = \text{constant}$ , modulo les problèmes discutés dans la section 5.2.1. Cependant, cette méthode est en général discutable pour des relations  $M/L$  non triviales comme dans l'équation (5.2) ou si une valeur différente de  $M/L$  est choisie pour différents types de galaxies. Dans ces deux derniers cas, la correction de la masse manquante devrait être appliquée à la distribution de masse elle-même au lieu de le faire sur celle de la luminosité, afin d'éviter des erreurs systématiques sur l'affectation des masses, et par conséquent sur les vitesses reconstruites. Malheureusement, ceci nécessite une hypothèse préliminaire sur la valeur de  $\Omega_m$ , mais en définitive ne complique l'analyse que légèrement.

Dans les données observationnelles, les galaxies sont séparées en deux populations : les groupes<sup>2</sup> de galaxies (Tully, 1987) et les galaxies de champ. Ces deux populations doivent être traitées séparément, en gardant à l'esprit que les groupes sont les plus critiques car leur influence gravitationnelle est plus grande que les galaxies de champ individuelles et de plus leur vitesses propres mesurées est plus précise. Nous allons à présent détailler la procédure complète consistant à créer un catalogue virtuel limité en magnitude et à "reconstruire" la distribution de masse. Nous rappelons que, dans nos catalogues virtuels, les groupes de galaxies simulés sont les halos de matière noire de plus de 5 particules alors que les galaxies de champ sont identifiées avec les particules de matière noire qui n'ont été associées à aucun halo. Nous allons au préalable présenter les principales étapes de la correction de l'incomplétude :

- I. Les luminosités totales apparentes des groupes de galaxies sont obtenues en supposant soit une fonction de luminosité Schechter globale ou locale (suivant les informations disponibles) pour les groupes considérés. La luminosité intrinsèque est calculée trivialement à partir de la luminosité apparente et le redshift du groupe.
- II. La luminosité intrinsèque des galaxies qui n'ont pas été liées (donc des galaxies de champ) est aussi déterminée d'une manière directe.
- III. Ensuite, les masses sont estimées en affectant le rapport  $M/L$  adéquat à chaque objet de I et de II.
- IV. La masse manquante locale qui est due aux galaxies de champ non détectées est déduite de la distribution de masse observée. Ceci nécessite un a priori sur  $\Omega_m$ .
- V. Cette masse manquante peut-être soit réaffectée localement aux galaxies de champ du II (ce que nous avons choisi) ou être introduite par le biais de traceurs placés aléatoirement, comme dans la section 5.2.1.

A présent, nous pouvons étudier la procédure en détail.

<sup>2</sup>Les groupes sont définis comme des ensembles compacts de 5 galaxies ou plus.

### Simulation et correction des effets d'incomplétude dans les catalogues de galaxie

N'ayant à notre disposition qu'un catalogue de halos, nous devons fabriquer un "catalogue de galaxies" avec des effets d'incomplétude. La principale difficulté dans cette construction est que la distribution des galaxies dans l'Univers est une fonctionnelle non-triviale et non-linéaire du champ de densité de la matière. Par exemple, les galaxies brillantes ont tendance à se concentrer dans les structures massives (Zandivarez et al., 2006). Cela signifie que, bien qu'il nous manque la plupart des galaxies de champ, les grands groupes peuvent encore être vus à cause des galaxies brillantes qu'ils contiennent. Par conséquent, la distribution des galaxies devrait principalement tracer les grands halos à grande distance, ce qui peut introduire un biais potentiel dans la reconstruction des vitesses propres si les corrections d'incomplétude sont effectuées sans trop de précautions.

Dans ce qui suit, nous générons les catalogues virtuels de galaxies qui ressemblent au NBG-8k/3k. Pour tenir compte correctement des effets qui ont été discutés ci-dessus, nous devons séparer les groupes de galaxies des galaxies de champ. Les groupes sont peuplés avec des galaxies en utilisant une forme de Schechter universelle pour rester simple, mais avec une normalisation différente pour tenir compte de leur  $M/L$  non-trivial.

Statistiquement, les catalogues NBG-8k/3k sont composé de galaxies observées en bande B et distribuées d'après la forme de Schechter

$$n(L) dL \simeq n_0 L^{-1} e^{-L/L_*} dL, \quad (5.6)$$

avec  $L_* \simeq 5.7 \times 10^{10} L_\odot$  et  $n_0 \simeq 0.03 h^3 \text{Mpc}^{-3}$ . De plus, le catalogue NBG-8k est complet pour des luminosités au-dessus de  $3 \times 10^9 - 4 \times 10^9 L_\odot$  à l'intérieur d'une sphère de rayon  $d_{\text{comp}} = 12 h^{-1} \text{Mpc}$ . Comme la densité moyenne de "galaxies" (particules) dans la simulation est de  $n_{\text{sim}} = 0.26 h^3 \text{Mpc}^{-3}$  et  $n_{\text{cat}} = 0.08 h^3 \text{Mpc}^{-3} \simeq 0.30 n_{\text{sim}}$  dans NBG-8k, nous devons diluer la simulation pour obtenir un catalogue virtuel semblable au NBG-8k. La luminosité  $L_G$  d'une galaxie détectée à une distance  $d$  de l'observateur doit satisfaire la contrainte

$$L_G > 4\pi l_{\text{cut}} d^2 \quad (5.7)$$

avec  $l_{\text{cut}}$  le flux minimal détectable par l'observateur. La fraction de galaxies détectée à la distance  $d$  dans le catalogue virtuel de galaxies est donc

$$f_{\text{field}}(d) = \begin{cases} 0.30 & \text{if } d < d_{\text{comp}} \\ \frac{\int_{4\pi l_{\text{cut}} d^2}^{\infty} n(L) dL}{\int_{L_{\text{min}}}^{\infty} n(L) dL} & \text{sinon} \end{cases} \quad (5.8)$$

avec  $l_{\text{cut}}$  le flux minimum détectable par l'observateur. La fraction est saturée à 0.30 pour satisfaire la contrainte de dilution exprimée plus haut. Nous renforçons la continuité de  $f_{\text{field}}(d)$  en choisissant  $L_{\text{min}}$  tel que  $f_{\text{field}}(d_{\text{comp}}) = 0.30$ . Le catalogue virtuel de galaxies et de groupes de galaxies est maintenant construit :

- I. Nous prenons un halo  $A$  de *FullMock* et nous supposons que c'est un groupe de galaxies. Nous déduisons donc sa luminosité intrinsèque  $L_A$  de sa masse  $M_A$  en utilisant l'équation. (5.2).
- II. La luminosité observé  $L'_A$  de  $A$  est calculée en supposant que sa population de galaxie suit (5.6) mais avec une normalisation différente afin d'obtenir la luminosité intrinsèque  $L_A$ . Si  $d_A$  est la distance entre l'observateur et le halo  $A$ , alors les galaxies détectées dans ce halo vérifient (5.7) pour  $d = d_A$ . La luminosité totale observable pour  $A$  est donc

$$L'_A = L_A f_L(d_A) \quad (5.9)$$

avec, en supposant  $L_{\text{min}} \ll L_*$ ,

$$f_L(d) = \begin{cases} \sim 1 & d < d_{\text{comp}} \\ \frac{\int_{4\pi d^2 l_{\text{cut}}}^{\infty} L n(L) dL}{\int_{L_{\text{min}}}^{\infty} L n(L) dL} \simeq e^{-\frac{4\pi d^2 l_{\text{cut}}}{L_*}} & d \geq d_{\text{comp}} \end{cases} \quad (5.10)$$

- III. Si  $L'_A < 4\pi d_A^2 l_{\text{cut}}$  alors  $A$  est supprimé du catalogue, sinon il est gardé.
- IV. Cela nous donne la composante “groupe” de notre catalogue limité en magnitude.
- V. Le cas des “galaxies de champ” est traité séparément. Les galaxies sont identifiées aux particules de matières noires et leur luminosité est calculée en supposant (5.6). De manière plus détaillée, nous choisissons une coquille  $S_d$  mise à une distance  $d$  de l’observateur. La probabilité de garder une “galaxie”  $G$  dans  $S_d$  est donnée par (5.8). A l’intérieur de la coquille  $S_d$ , les galaxies sélectionnées partagent maintenant une luminosité

$$L_f(d) = \int_{4\pi d_G^2 l_{\text{cut}}}^{\infty} L n(L) dL \quad (5.11)$$

qui est redistribuée également entre elles. Strictement parlant, une telle répartition devrait être effectuée aléatoirement d’après (5.6). Cela ajouterait un petit bruit additionnel sur les vitesses reconstruites. Ce bruit ne devrait pas avoir de conséquences significatives d’après la discussion du cas  $TS-T$  dans la section § 5.2.2.

Nous avons à présent un catalogue virtuel réaliste et nous allons essayer de tenir compte des effets d’incomplétude comme nous le ferions pour NBG-8k :

- A. La luminosité manquante dans les groupes est corrigée. Pour cela, nous calculons, dans une coquille mince  $S_d$  à une certaine distance  $d$ , le rapport entre la luminosité totale attendue et la luminosité observée

$$b(d) = \frac{\int_0^{\infty} L n(L) dL}{\int_{4\pi d^2 l_{\text{cut}}}^{\infty} L n(L) dL} = e^{4\pi d^2 l_{\text{cut}}/L_*}. \quad (5.12)$$

La luminosité intrinsèque  $L_A$  d’un groupe  $A$  dans  $S_d$  est reconstruite grâce à l’équation

$$L_A = L_{\text{obs},A} b(d). \quad (5.13)$$

La masse  $M_A$  de  $A$  peut maintenant être obtenue avec la relation *non-linéaire* (5.2).

- B. Le restant de la masse manquante dans  $S_d$  peut maintenant être écrite comme

$$M_{\text{missed},d} = \Upsilon b(d) (L_{\text{field,obs},d} + L_{\text{group,obs},d}) - M_{\text{field,obs},d} - M_{\text{group,obs},d}, \quad (5.14)$$

avec  $\Upsilon = 93 \frac{M_{\odot}}{L_{\odot}}$  le rapport  $M/L$  moyen,<sup>3</sup>  $L_{\text{group,obs},d}$  la luminosité observée des groupes,  $M_{\text{group,obs},d}$  les masses des groupes obtenues après la correction ci-dessus,  $L_{\text{field,obs},d}$  la luminosité des galaxies de champ. La quantité  $M_{\text{missed},d}$  vient à la fois des galaxies et des groupes manquants de  $S_d$ . Si  $M_{\text{missed},d} > 0$  et sans plus d’information, la masse manquante peut soit être affecté de manière uniforme aux galaxies de champ de  $S_d$  (ce que nous avons choisi et qui correspond à ce qui est généralement fait dans la littérature), ou distribué uniformément dans  $S_d$  en utilisant de nouveaux traceurs virtuels. Si  $M_{\text{missed},d} \leq 0$ , la distribution de masse dans  $S_d$  n’est pas touchée. Cette procédure comporte très probablement des biais. Par exemple, le contraste entre des coquilles est partiellement lissé, comme cela est illustré par la graphique de la figure 5.6. On pourrait dire que c’est quasiment équivalent à réduire l’amplitude typique des fluctuations dans le champ de densité. Par conséquent, un petit biais vers un  $\Omega_m$  plus grand peut advenir, comme le montre le panneau en bas à droite de la figure 5.2. Au contraire, si la masse manquante est affectée à de nouveaux traceurs virtuels, on s’attend à ce que le  $\Omega_m$  estimé soit inférieur à la vraie valeur comme il a été discuté dans la section 5.2.1.

- C. A présent la masse d’une “galaxie de champ” n’est plus la masse d’une particule. La procédure expliquée dans la section 3.3.2 est transformée comme suit, afin de garder l’algorithme simple. Avec  $v \in (0; 1]$  une variable aléatoire uniforme, une galaxie  $G$  de masse  $m_G$  est divisée en  $n_G$  sous-composantes de masse  $m_{\text{particle}}$  telles que :

$$n_G = \begin{cases} r_G & \text{si } \left( \frac{m_G}{m_{\text{particle}}} - r_G \right) < v \\ r_G + 1 & \text{autrement} \end{cases}, \quad (5.15)$$

<sup>3</sup>Il faut remarquer ici qu’un a-priori sur la valeur de  $\Omega_m$  est nécessaire pour estimer  $\Upsilon$ .

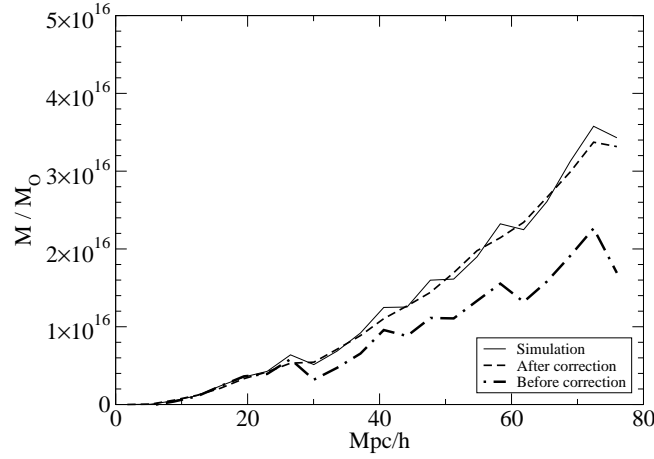


FIG. 5.6: *Limitation en magnitude/Correction de la masse manquante* – Ce graphe donne la quantité mesurée de masse dans une coquille fine à différentes distances de l’observateur. La ligne solide donne la distribution originale de masse dans la simulation, la ligne en traits-pointillés donne la distribution de masse après avoir simulé l’incomplétude et la ligne en traits interrompus donne la distribution de masse calculée après correction de l’incomplétude comme cela est décrit dans la section 5.2.3.

avec  $r_G = \left\lfloor \frac{m_G}{m_{\text{particle}}} \right\rfloor$ ,  $\lfloor x \rfloor$  étant la partie entière de  $x$ . Chacune de ces sous-composantes est maintenant considérée comme une “galaxie de champ” dans la procédure expliquée dans la section 5.2.3.

### Analyse

Pour étudier les effets systématiques dans la correction de l’incomplétude, nous avons utilisé *8k-mock6* et choisi une limite de flux telle que le catalogue virtuel corresponde à une incomplétude semblable au NBG-8k, comme cela est montré par le graphique de la figure 5.7. Les résultats sont résumés sur la figure 5.8 et dans la table 5.2.

Les vitesses propres radiales reconstruites  $v_{r,\text{rec}}$  se comportent extrêmement bien. En moyenne, la comparaison entre les champs de vitesse reconstruits et simulés est étonnamment bonne sur un volume de  $80 h^{-1}\text{Mpc}$ , même si au bord il nous manque 98% des galaxies de champ, qui représentent 60% de la masse totale dans notre catalogue virtuel. Cela signifie que, bien que nous ne gardions que 2% des galaxies de champ, elles suffisent, avec les groupes, pour avoir une bonne reconstruction du champ de vitesse à grande échelle. On remarque un petit biais dans le diagramme de dispersion du panneau en bas à droite de la figure 5.8, ce qui produit un  $\Omega_m = 0.38$  légèrement plus grand que la valeur attendue de 0.30, mais en bon accord avec la valeur effective de 0.35, qui est attendue dans ce volume (cf. section 5.4.3 sur les effets de variance cosmique). Ce biais peut être la conséquence de notre traitement de la masse manquante provenant de traceurs non détectés comme il a été discuté en détail ci-dessus (point B).

## 5.3 Distortion de décalage vers le rouge (redshift)

Les données d’entrée pour la reconstruction MAK sont les positions des objets dans l’espace réel, sans distortions en redshift, comme l’exprime l’équation (2.100). Cependant, les catalogues en redshift nous donnent les positions des galaxies dans l’espace des redshifts, c’est-à-dire  $s_r = Hd + v_r$ , où  $s_r$  est la distance en redshift,  $d$  est la distance lumineuse entre l’observateur et l’objet et  $v_r$  est la composante sur la ligne de visée de la vitesse propre de ce même objet. Pour tenir compte des distortions qui en découlent, nous devons corriger deux effets majeurs :



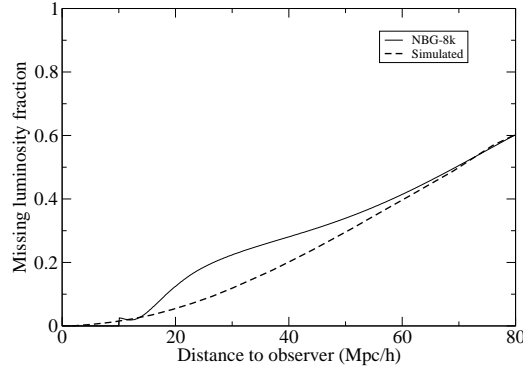


FIG. 5.7: *Limitation en magnitude/Fraction de luminosité manquante* – Ligne solide : Incomplétude prévue pour la luminosité du catalogue NBG-8k à une distance donnée de l’observateur. Ligne en trait interrompu : Incomplétude simulée dans *8k-mock6* L’incomplétude est exprimée en terme de fraction manquante de la luminosité à une distance donnée.

TAB. 5.2: *Incomplétude : limitation en magnitude* – La description des colonnes est donnée dans la légende la table 5.1 (page 74).

Volume	Velocity field			$\Omega_m (s)$	$\Omega_m (\mathcal{L}_{\min})$	$\Omega_m (\mathcal{L}_{\max})$	$\Omega_m (1.5\sigma, s_{\text{med}})$	$\Omega_m (1.5\sigma, s_{\min})$	$\Omega_m (1.5\sigma, s_{\max})$
	$s$	$r$	$\sigma$						
8k	0.86	0.77	0.64	0.39	0.26	0.31	0.29	0.25	0.34
4k	0.77	0.75	0.66	0.48	0.37	0.45	0.38	0.30	0.47

- Les “Fingers-of-god” (littéralement doigts de dieu) correspondent à une élongation des structures denses le long de la ligne de visée, telles que les amas de galaxies, à cause des mouvements aléatoires des galaxies dans ces structures (cf. section 6.2.3 pour une plus longue discussion)
- L’effet Kaiser (Kaiser, 1987), qui est un effet à grande échelle dû aux flots cosmiques, qui, par exemple augmente le contraste de densité global.

Les effets de “Fingers-of-god” peuvent être assez facilement corrigés en effondrant les groupes ou les amas en un point, comme cela est généralement effectué dans la littérature. Cependant, une telle procédure est d’une part généralement ad-hoc et très probablement entachée de biais d’autre part (cf. section 6.2.3).

On peut tenir compte de l’effet Kaiser en modifiant la fonction coût (2.100) en utilisant l’approximation de Zel’dovich (2.64) pour en déduire les vitesses selon la ligne de visée comme une fonction du champ de déplacement recherché (Mohayaee & Tully, 2005; Valentine et al., 2000). Si  $\mathbf{s}(\mathbf{q})$  est la coordonnée redshift d’une particule originellement à  $\mathbf{q}$  alors :

$$\mathbf{s} = \mathbf{x} + \mathbf{v}(\mathbf{q}) \cdot \hat{\mathbf{s}}, \quad (5.16)$$

TAB. 5.3: *Reconstruction en redshift* – La description des colonnes est donnée dans la légende de la table 5.1 (page 74).

$s$	$r$	$\sigma$	$\Omega_m (s)$	$\Omega_m (\mathcal{L}_{\min})$	$\Omega_m (\mathcal{L}_{\max})$	$\Omega_m (1.5\sigma, s_{\text{med}})$	$\Omega_m (1.5\sigma, s_{\min})$	$\Omega_m (1.5\sigma, s_{\max})$
0.83	0.46	0.95	0.50	0.22	0.29	0.27	0.22	0.33

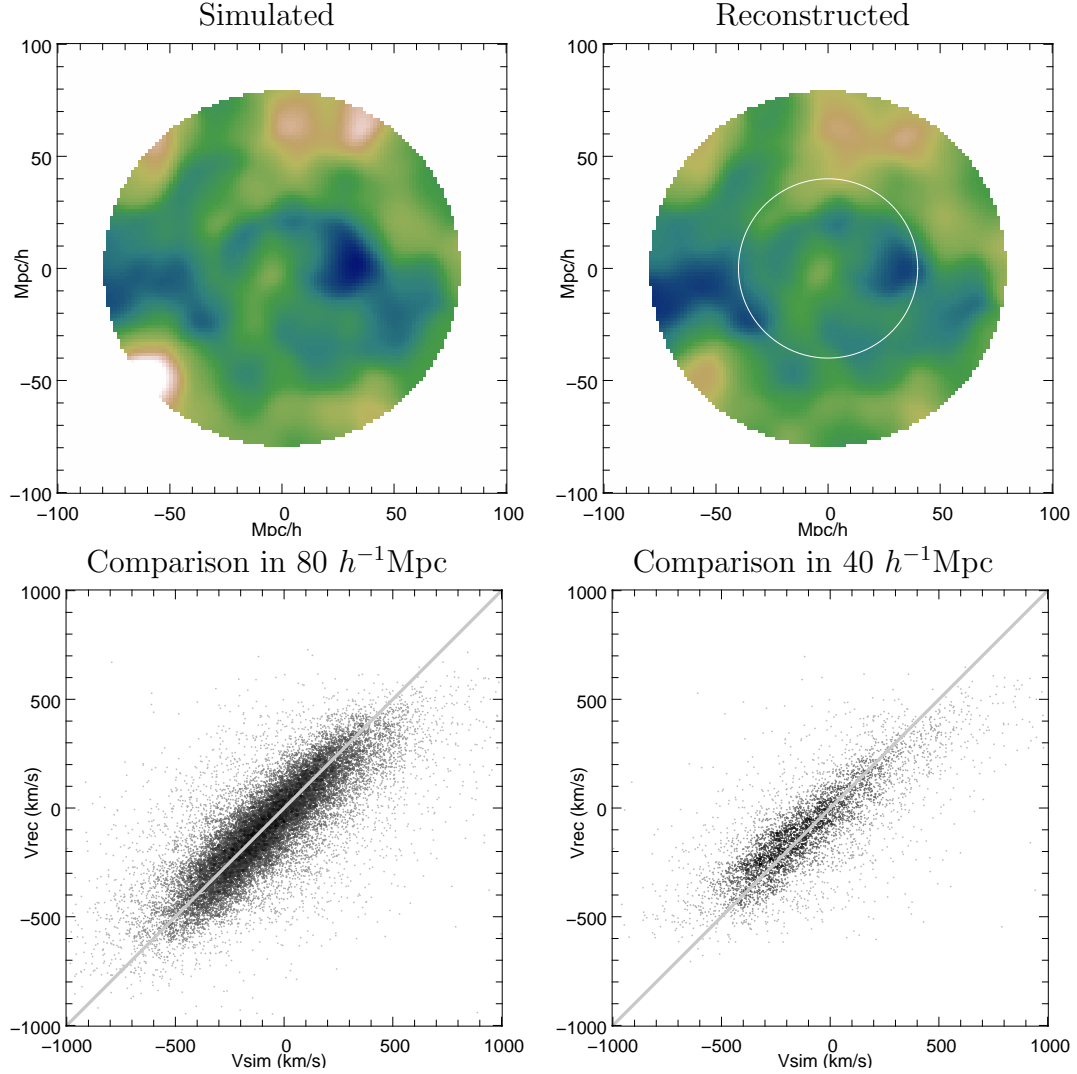


FIG. 5.8: *Incomplétude : limitation en magnitude* – Panneaux du haut : une coupe de la composante selon la ligne de visée du champ de vitesse simulé dans 8k-mock6 et celui reconstruit, après lissage avec une fenêtre gaussienne de  $5 h^{-1} \text{Mpc}$ . La coupe affichée est choisie afin d’inclure l’observateur en  $(0, 0)$ . Le cercle blanc dans le panneau de droite donne la taille de la sphère de  $40 h^{-1} \text{Mpc}$  inclus dans une de  $80 h^{-1} \text{Mpc}$ . Panneaux du bas : le diagramme en dispersion compare les vitesses reconstruites et simulées des objets dans la région de  $80 h^{-1} \text{Mpc}$  de rayon (panneau de gauche) et le volume de  $40 h^{-1} \text{Mpc}$  (panneau de droite).

avec  $\hat{\mathbf{s}}$  le vecteur ligne de visée de la particule,

$$\mathbf{v}(\mathbf{q}) = \beta(\mathbf{x}(\mathbf{q}) - \mathbf{q}) , \quad (5.17)$$

avec  $\beta$  la facteur de croissance linéaire et  $H$  la constante de Hubble. Dans ces expressions, on exprime les déplacements et les positions en unités de la vitesse ( $\text{km s}^{-1}$ ), ce qui permet d'omettre la constante de Hubble dans toutes les expressions. Le déplacement réel va donc être

$$\mathbf{x}(\mathbf{q}) - \mathbf{q} = \mathbf{s}(\mathbf{q}) - \mathbf{q} - \beta((\mathbf{x}(\mathbf{q}) - \mathbf{q}) \cdot \hat{\mathbf{s}}) \hat{\mathbf{s}} , \quad (5.18)$$

soit en projection sur la ligne de visée :

$$(\mathbf{x}(\mathbf{q}) - \mathbf{q}) \cdot \hat{\mathbf{s}} = \frac{1}{1 + \beta} (\mathbf{s}(\mathbf{q}) - \mathbf{q}) \cdot \hat{\mathbf{s}} . \quad (5.19)$$

Maintenant, on peut réécrire le déplacement (5.18) uniquement en terme de  $\mathbf{s} - \mathbf{q}$  :

$$\mathbf{x}(\mathbf{q}) - \mathbf{q} = \mathbf{s}(\mathbf{q}) - \mathbf{q} - \frac{\beta}{1 + \beta} ((\mathbf{s}(\mathbf{q}) - \mathbf{q}) \cdot \hat{\mathbf{s}}) \hat{\mathbf{s}} . \quad (5.20)$$

En prenant cette expression au carré et en remplaçant dans l'équation (2.100), on obtient la nouvelle fonction coût de l'application  $\sigma$

$$I_\sigma = \sum_{i=1}^N \left( (\mathbf{s}_i - \mathbf{q}_{\sigma(i)})^2 - \frac{\beta(2 + \beta)}{(1 + \beta)^2} \frac{((\mathbf{s}_i - \mathbf{q}_{\sigma(i)}) \cdot \mathbf{s})^2}{\|\mathbf{s}\|^2} \right) . \quad (5.21)$$

Une fois que le déplacement  $\Psi^s = \mathbf{s} - \mathbf{q}$  a été calculé, la vitesse propre reconstruite de l'objet  $i$  peut être obtenue avec l'égalité suivante :

$$v_{\text{r,rec}}^s = \frac{\beta}{1 + \beta} \frac{\mathbf{s} \cdot \Psi^s}{\|\mathbf{s}\|} . \quad (5.22)$$

La fonction coût  $I_\sigma$  nous amènerait au résultat exact pour un champ de déplacement dans l'approximation de Zel'dovich sans croisement de coquille dû aux distortions en redshift. Cependant, en général, le deuxième terme (qui tient compte de la distortion en redshift) de l'équation (5.21) devient du même ordre que le premier terme (le terme de coût dans l'espace réel) quand on est proche de l'observateur. Dans ce cas, la reconstruction devient mal définie à cause de la perte de convexité de la fonctionnelle  $I_\sigma$ . Nous nous attendons donc à ce que la partie centrale de tous les catalogues soit, en général, mal reconstruite. La taille d'une telle région est grossièrement déterminée par la magnitude  $v_{\text{obs}}$  d'un flot à grande échelle proche de l'observateur par rapport au fond diffus cosmologique. La vitesse  $v_{\text{obs}}$  détermine la contribution relative du premier terme par rapport au second terme de l'équation (5.21). En pratique,  $v_{\text{obs}}$  est de l'ordre de quelques centaines de  $\text{km s}^{-1}$  (par exemple, la vitesse du Groupe Local est  $630 \text{ km s}^{-1}$ , Erdoğdu et al., 2006b), ce qui nous donne une région d'exclusion d'un rayon de quelques  $h^{-1} \text{Mpc}$ .<sup>4</sup>

Encore une fois, la reconstruction MAK échoue dans les régions avec croisements de coquille. La projection en espace des redshift engendre de tels croisements le long de la ligne de visée. Ces croisements de coquilles sont dramatiques à cause de leur nature anisotrope. En particulier, les filaments peuvent se croiser en passant de l'espace réel à l'espace redshift, ce qui implique que la reconstruction échouera dans de grandes régions du catalogue qui sont sous l'influence gravitationnelle de ces filaments. Dans ces régions, la plupart des vitesses radiales reconstruites auront le signe opposé à celui de la vitesse vraie. Bien sur, les croisements de coquilles en coordonnées redshift peuvent avoir des conséquences plus complexes, mais ce simple exemple suggère que la reconstruction MAK ne devrait pas aussi bien fonctionner en espace redshift qu'en espace réel.<sup>5</sup>

<sup>4</sup>Voir, par exemple, Colombi et al. (2007) pour une discussion semblable.

<sup>5</sup>Ceci est aussi vrai pour la méthode de moindre action pour laquelle on obtient de multiples solutions.

Un autre problème inhérent à cette méthode est lié à l'hypothèse faite sur la valeur de  $\beta$  avant la reconstruction. Comme pour la section 5.2.3, où nous devons deviner la masse non détectée, nous choisissons une valeur  $\Omega_{m,in}$ , et donc un  $\beta_{in}$  supposé, puis nous faisons une reconstruction redshift et mesurons un  $\Omega_{m,out}$ . En pratique, le “vrai”  $\Omega_m$  du catalogue a été choisi comme celui pour lequel  $\Omega_{m,in} = \Omega_{m,out}$ , qui correspond à avoir une modélisation des orbites cohérente à la fois quand on effectue la reconstruction MAK et quand on les compare aux vitesses observées.

La figure 5.9 montre à la fois les champs de vitesse reconstruits et simulés et le diagramme de dispersion entre  $v_{r,rec}^s$  et  $v_{r,sim}$ . La première impression lors de la comparaison des deux panneaux du haut de la figure 5.9 est que la reconstruction redshift marche vraiment bien. Cependant, il existe quelques points potentiellement inquiétants :

- Des structures importantes ont de mauvaises vitesses reconstruites. Deux exemples importants sont le doigt jaune-vert situé juste au-dessus du centre du panneau du haut de la figure 5.9 et le gros pic de vitesse en haut du même panneau. Dans le panneau de gauche, ces deux structures ne sont pas si proéminentes. La différence peut être comprise en étudiant l'impact de l'effet Kaiser sur le champ de vitesse reconstruit. Essentiellement, deux filaments voisins peuvent fusionner dans l'espace redshift et donner lieu à un filament avec une densité apparente plus grande. La reconstruction n'est pas capable de séparer ces deux filaments, ce qui conduit à une région de vitesses reconstruites plus élevées que les vraies. Donc, nous nous attendons à ce que les données observationnelles aient les mêmes problèmes dans le voisinage du “Great Wall”, qui est un superamas de filaments comprimés par la distortion redshift.
- Le champ de vitesse dans le voisinage immédiat ( $5-10 h^{-1}\text{Mpc}$ ) de l'observateur virtuel n'a pas la bonne structure spatiale, comparée à celle du champ simulé, et présente même un pic excédentaire. C'est malheureusement un problème auquel on s'attendait et qui est lié à la discussion ci-dessus sur les défauts de  $I_\sigma$  pour les particules proches de l'observateur. En effet, dans le voisinage de l'observateur,  $I_\sigma$  devient singulier et la reconstruction concernant les vraies orbites est très probablement erronée. Une analyse du champ de vitesse lissé semble montrer que cet effet ressemble en pratique à celui qui a été présenté dans le point précédent : le champ de vitesse reconstruit peut être amplifié par la fusion de différentes structures dans le voisinage de l'observateur.
- Le panneau en bas à droite présente deux queues anti-diagonales supplémentaires comparé à celui en bas à gauche. Comme discuté plus tôt, ces queues sont dues à des croisements de coquille se produisant le long de la ligne de visée quand on passe des coordonnées réelles aux coordonnées redshifts. Ces croisements de coquilles supplémentaires résultent du fait que certaines vitesses reconstruites acquièrent le mauvais signe par rapport aux vitesses vraies.

Comme dans la section 4.1.1, nous avons calculé dans la figure 5.10 la distribution des différences  $P_{VE}^s$  entre  $v_{r,rec}^s$  et  $v_{r,sim}$ , pour une reconstruction redshift appliquée à 8k-mock6 et basée sur un maillage  $64^3$ .<sup>6</sup> Bien que la distribution soit bien sur plus large que dans la figure 4.2, les conclusions précédentes sont encore valides.  $P_{VE}^s$  est mieux ajusté par une distribution Lorentzienne avec  $B = 86 \text{ km s}^{-1}$  que par une Gaussienne de largeur  $\sigma = 91 \text{ km s}^{-1}$ , particulièrement dans les queues de distribution.

Pour vérifier les effets dus aux distortions en redshift sur la qualité de la reconstruction, on peut comparer la table 5.3 à la première ligne de la table 5.1. Comme d'habitude, le paramètre  $s$  est légèrement inférieur à 1 à cause des effets non-linéaires discutés dans la section 4.1.1, ce qui semble être légèrement amplifié par les distortions en redshift. L'apparition des queues anti-diagonales dans le panneau en bas à droite de la figure 5.9 augmente le degré de dispersion, ce qui fait décroître le coefficient de corrélation  $r$  et le rapport signal-sur-bruit augmente. Réduire l'analyse à la région à l'intérieur de l'isocontour à  $1.5\sigma$  améliore beaucoup les résultats, comme on s'y attend, mais amène tout de même à une valeur de  $\Omega_m$  légèrement biaisée vers des valeurs plus basses,  $\Omega_m = 0.27$ .

<sup>6</sup>La gestion des effets de volume fini est analysé dans la section 5.4.2.

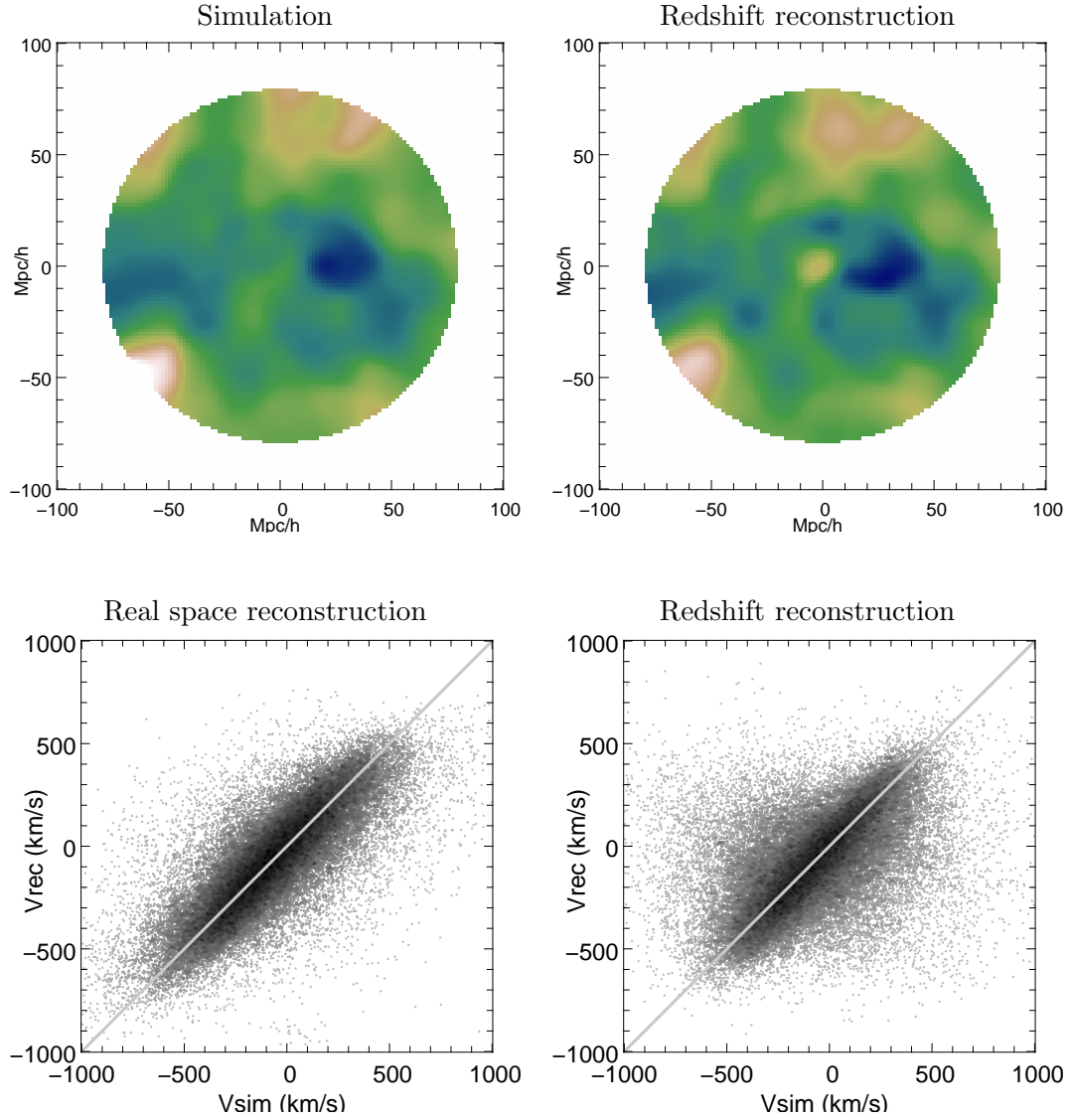


FIG. 5.9: *Correction de l'effet de distortion de décalage en rouge* – Panneaux du haut : Une coupe de la composante selon la ligne de visée des champ de vitesse lissés  $v_{r,\text{sim}}$  et  $v_{r,\text{rec}}$  est affichée respectivement dans les panneaux de gauche et de droite. Les deux champs ont été lissés avec une fenêtre gaussienne de  $5 h^{-1}\text{Mpc}$ , les objets ont été placés à leurs coordonnées comobiles réelles (simulées et reconstruites). Panneaux du bas : diagramme de dispersion entre  $v_{r,\text{sim}}$  et  $v_{r,\text{rec}}$  pour des traceurs de masse individuels. Le panneau de gauche (resp. de droite) a été produit en utilisant une reconstruction en espace réel (resp. reconstruction en espace redshift). Dans les deux cas, seuls les objets à l'intérieur d'une sphère de  $8000 \text{ km s}^{-1}$  ont été montrés.

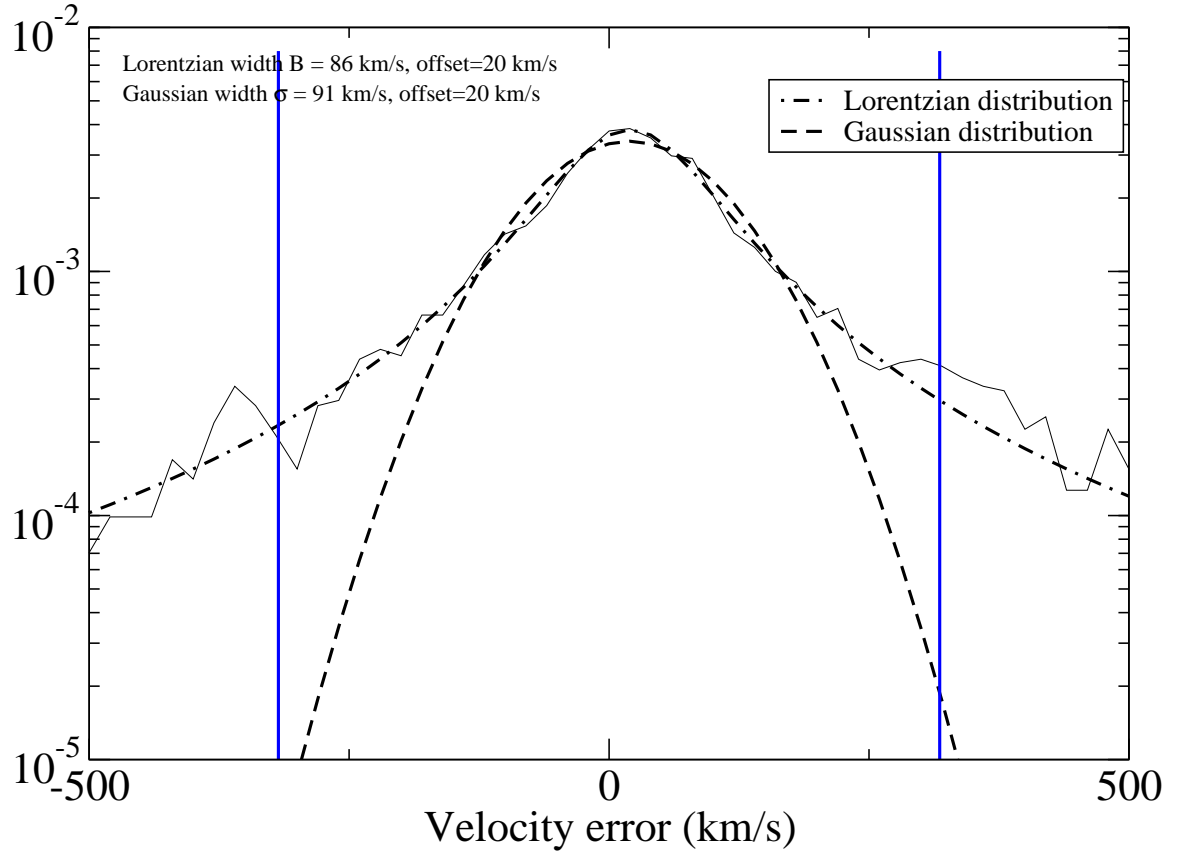


FIG. 5.10: *Distribution des erreurs du champ de vitesse reconstruit, espace redshift* – Même chose que pour la figure 4.2 (page 54) mais la ligne solide correspond à la distribution de probabilité de la quantité  $v_{r,\text{rec}}^s - v_{r,\text{sim}}$ , où  $v_{r,\text{rec}}^s$  et  $v_{r,\text{sim}}$  sont respectivement les vitesses reconstruites et simulées selon la ligne de visée.

## 5.4 Effets dues à la géométrie du catalogue

En pratique, les vrais catalogues de galaxies ne sont pas périodiques spatialement comme dans notre simulation. Ils représentent une région de volume fini avec une géométrie non-triviale. En particulier, deux types de problème peuvent subvenir :

- *Les effets de bord* – La reconstruction des trajectoires de galaxies sans avoir la moindre information sur ce qui peut dynamiquement les affecter de l’extérieur du catalogue peut très probablement introduire des erreurs significatives, et potentiellement systématiques. Nous séparons ici les effets de bord en deux sous-classes : les effets d’obscurcissement par notre galaxie, qui définit la “Zone Of Avoidance” (par la suite ZOA) et les effets de profondeur finie dans le catalogue. Ces deux effets recevront des traitements différents, qui sont détaillées dans les section 5.4.1 et 5.4.2.
- *La variance cosmique* – Le volume fini de la partie accessible de l’Univers peut être potentiellement une réalisation moins probable du processus aléatoire sous-tendant les propriétés à grande échelle de la distribution de matière. Nous devons savoir si notre méthode, en incluant les effets de bords, est robuste et capable de retrouver les propriétés statistiques de l’Univers entier à partir d’observations concernant seulement une petite fraction de ce dernier.

### 5.4.1 Le plan galactique de la voie lactée

La lumière est fortement absorbée par les poussières présentes dans la plan galactique de la voie lactée. Par conséquent, les catalogues de galaxies ne fournissent généralement pas de données dans la direction (approximativement la région pour laquelle  $|b| < 5$  deg, où  $b$  est la latitude galactique) de la ZOA. Cette absorption très forte introduit un effet de bord, qui a la caractéristique désagréable d’être présent à toute distance de l’observateur et peut donc fausser les mesures. Comme cette région est néanmoins relativement petite, tout particulièrement à bas redshift, une correction simple devrait être en mesure de supprimer une grande partie de cet effet de bord dans la région centrale du catalogue. Simuler l’effet est facile en introduisant un observateur au centre du volume de la simulation et en supprimant tous les traceurs de masse dans le voisinage du plan galactique  $z = 0$ , *e.g.* ceux qui ont  $|b| < \alpha$ . Cela nous donne le catalogue *FullMockZOA*.<sup>7</sup>

Bien que des façons plus avancées existent de remplir la ZOA (par exemple Lahav et al., 1994; Fontanot et al., 2003), cette région est ici suffisamment petite pour qu’on s’en occupe avec l’algorithme simple qui suit. Comme les propriétés statistiques des galaxies ne devraient pas changer en passant à travers le bord de la ZOA, les objets dans son voisinage peuvent être utilisés pour remplir le volume manquant. Nous construisons de nouveaux traceurs de masse virtuels pour remplir la région “obscurcie” en appliquant une transformation de symétrie localement planaire aux galaxies et groupes satisfaisant la condition  $-3\alpha < b < -\alpha$  par rapport au “plan”  $-\alpha$ . Nous exécutons la même opération sur les objets présents dans le volume décrit par  $+\alpha < b < +3\alpha$  mais pour une symétrie par rapport au “plan”  $+\alpha$  (cf. schéma explicatif de la figure 5.11). A la fin, les masses des halos qui ont été copiés dans la ZOA sont divisés par deux et nous ne prenons que la moitié des galaxies de champ. Cette méthode a été utilisée précédemment pour remplir la ZOA dans le catalogue NBG-3k (Shaya et al., 1995, voir aussi 6.1) et dans le NBG-8k. Cette procédure de repliement a été appliquée à *FullMockZOA*, en bougeant légèrement quelques-uns des nouveaux objets dans le but de récupérer la périodicité de la boîte de simulation pour éviter de mélanger des effets dues à la ZOA avec d’autres effets de bord. Les résultats sont présentés par les graphes de la figure 5.13. Comme on s’y attend, la ZOA a un impact net sur les vitesses reconstruites. La figure 5.12 illustre cette procédure dans le cas de 8k-mock6 qui est centré de la même manière que *FullMock* et *FullMockZOA*.

Les erreurs typiques sur les vitesses, représentées sur le panneau de gauche de cette figure, augmentent nettement dans le voisinage de la région obscurcie. Heureusement, elles restent bien en-dessous de la dispersion naturelle du champ de vitesse de la simulation (ligne en tiret). Comme nous comparons les champs de vitesse filtrés avec une fenêtre gaussienne de rayon  $5 h^{-1}\text{Mpc}$ , nous nous attendons

<sup>7</sup>  $\alpha = 5$  deg dans notre cas.

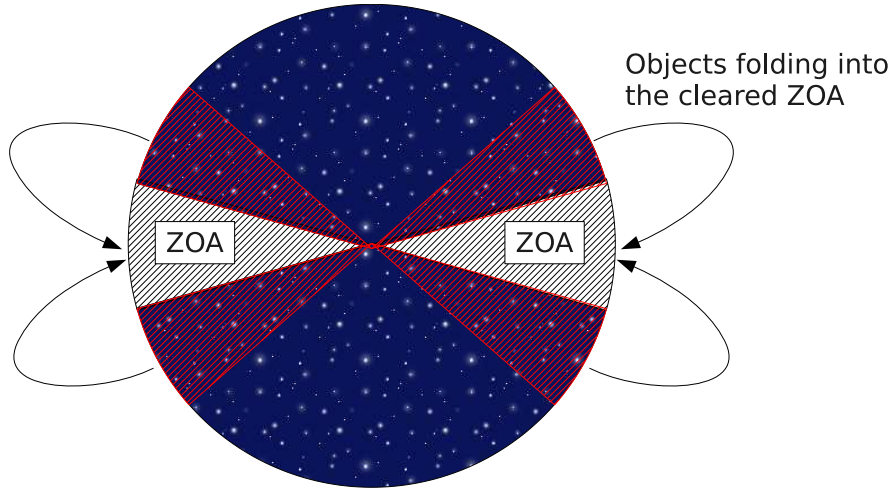


FIG. 5.11: *Illustration de l’algorithme de remplissage du plan galactique* – La zone hachurée en noir est vide de galaxie. On utilise les galaxies de la zone hachurée en rouge pour remplir la zone en hachures noires.

TAB. 5.4: *Obscurcissement du plan galactique* – Résumé des mesures de bruit et de biais. La description des colonnes est donné dans la légende de la table 5.1 (page 74).

$s$	$r$	$\sigma$	$\Omega_m$ ( $s$ )	$\Omega_m$ ( $\mathcal{L}_{\min}$ )	$\Omega_m$ ( $\mathcal{L}_{\max}$ )	$\Omega_m$ ( $1.5\sigma, s_{\text{med}}$ )	$\Omega_m$ ( $1.5\sigma, s_{\min}$ )	$\Omega_m$ ( $1.5\sigma, s_{\max}$ )
0.89	0.79	0.61	0.37	0.30	0.35	0.32	0.285	0.36

à ce que le *champ* de vitesse reconstruite soit quasiment sans erreur pour tous les points à moins de  $\sim 60h^{-1}\text{Mpc}$  de l’observateur.<sup>8</sup> Il est aussi heureux que nous n’ayons introduit aucun biais avec l’algorithme de remplissage, comme on peut le voir en comparant la table 5.4 à la première ligne de la table 5.1 et en regardant le diagramme de dispersion dans le panneau de droite de la figure 5.13. Néanmoins, il faut bien faire attention que les effets de bord ne soient pas du tout localisés au voisinage de la ZOA mais s’étendent assez loin et ne deviennent négligeables que pour  $|b| > 20$  deg. La table 5.4 montre que le bruit qui a été ajouté aux vitesses reconstruites lors de ce traitement n’a aucun effet sur le  $\Omega_m$  mesuré.

### 5.4.2 Le domaine lagrangien

Les données d’entrée de la reconstruction MAK sont les coordonnées des objets à l’instant présent, c’est-à-dire les  $\mathbf{x}$  dans l’équation (2.100) ou les  $\mathbf{s}$  dans l’équation (5.21), et la connaissance du domaine lagrangien,  $\mathbf{q}$  dans l’équation (2.100) ou (5.21). Les catalogues de redshift donnent les “positions” actuelles des objets, *i.e.* les  $\mathbf{s}$  de (5.21), mais nous n’avons aucune observation qui pourrait nous donner le domaine lagrangien correspondant  $\{\mathbf{q}\}$ . Nous sommes donc limités à faire des conjectures, bien qu’à la fin, pour des catalogues volumineux, les détails de cette conjecture n’ont pas d’importance car les forces gravitationnelles sont écrantées aux grandes échelles par la distribution quasi-homogène de matière dans l’Univers. Par conséquent, ce qui arrive au bord ne devrait pas affecter fortement la partie centrale du catalogue bien que certaines conjectures soient meilleures à confiner les effets de bords sur les parties externes du catalogue. Une solution naïve est de supposer que le domaine lagrangien n’est pas si différent du volume du catalogue lui-même. Cette hypothèse ne commence à

<sup>8</sup>Cela correspond à prendre une ZOA large de 5 deg et à calculer à quelle distance la fenêtre est plus petite que la ZOA.



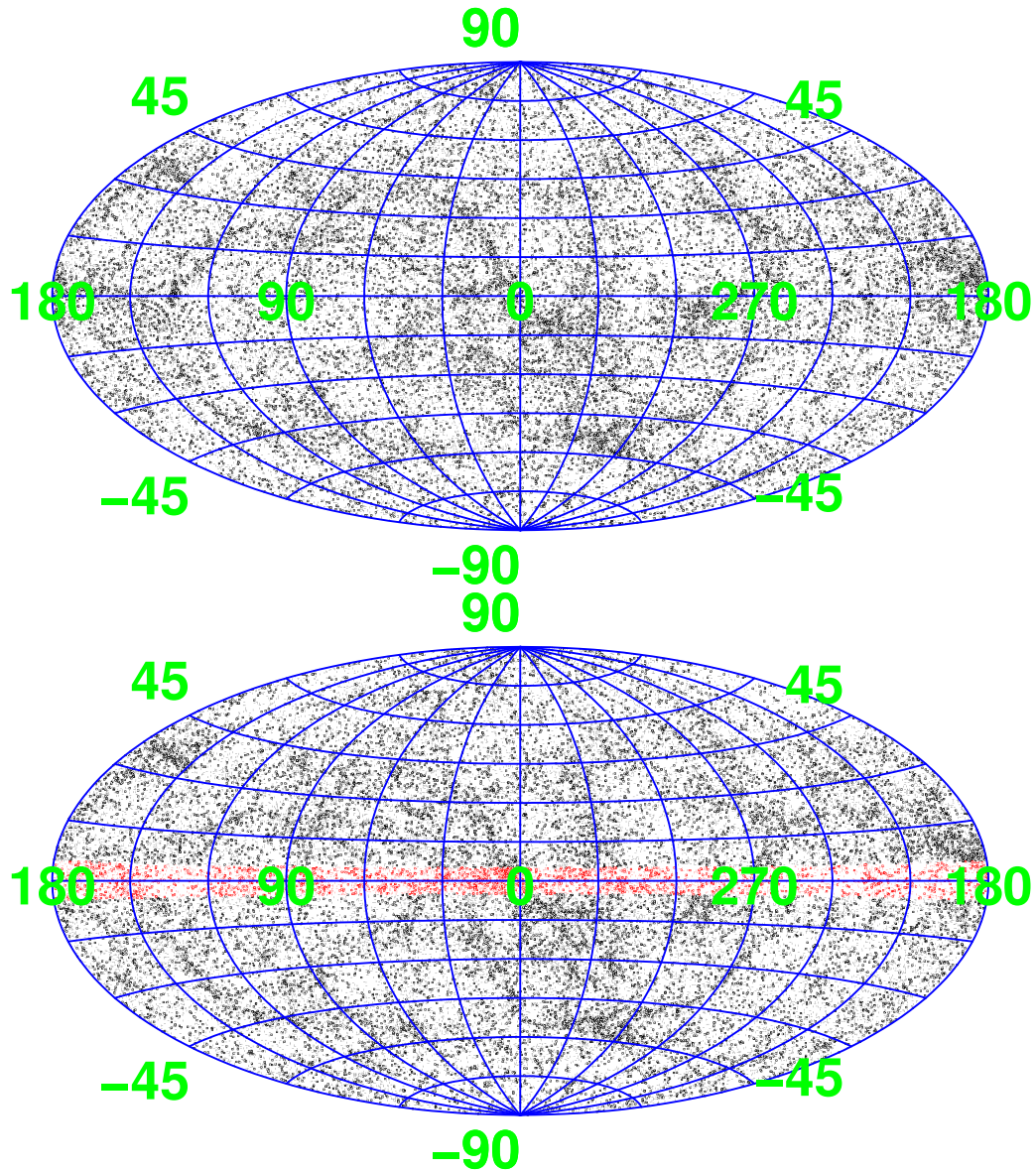


FIG. 5.12: *Simulation du plan galactique de la voie lactée* – On présente ici 8k-mock6 en projection aït-off tel qu’il serait vu par un observateur situé au centre de celui-ci. Chaque point représente un groupe de galaxies ou une galaxie. Un point est noir s’il correspond à un objet réellement existant dans la simulation. Il est rouge s’il a été ajouté suivant l’algorithme expliqué dans la section 5.4.1. Le diagramme du haut représente le catalogue originel, le diagramme du bas, le catalogue après simulation de l’effet de bord dû à la voie lactée. On a utilisé une coupure à 5 deg du plan galactique. On constate qu’un assez gros amas est caché dans le plan galactique.

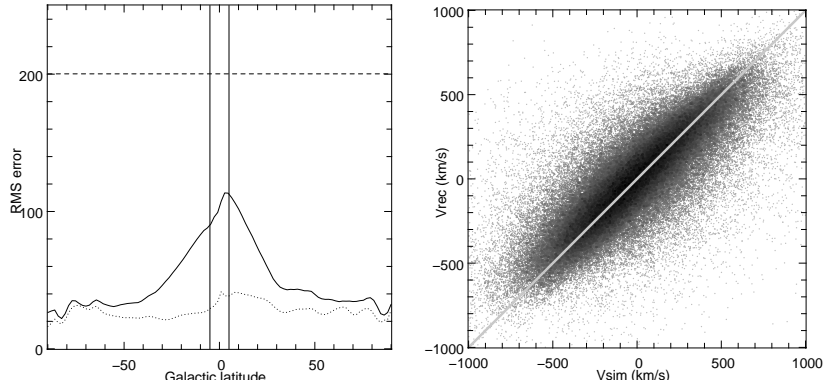


FIG. 5.13: *Obscurcissement du plan galactique / champ de vitesse* – Panneau de gauche : Histogramme de l’erreur moyenne RMS (Root Mean Square) du champ de vitesse lissé. Comme d’habitude, le champ de vitesse a été lissé avec une fenêtre gaussienne de  $5 h^{-1}\text{Mpc}$  de rayon. Chaque point est calculé en moyennant le carré de la déviation du champ de vitesse selon la ligne de visée et pour toutes les lignes de visée appartenant au même intervalle de  $\sin(b)$ , où  $b$  est la latitude “galactique”. La ligne pleine donne l’erreur RMS en présence de l’obscurcissement du plan galactique en  $b = 0$ . La ligne pointillée donne l’erreur RMS pour une reconstruction sur un catalogue sans obscurcissement. La ligne en traits interrompus donne l’erreur RMS du champ de vitesse lui-même. Panneau de droite : Dispersion entre  $v_{r,\text{rec}}$  et  $v_{r,\text{sim}}$  pour des traceurs individuels.

être une bonne approximation que pour des volumes inscrits dans une sphère de rayon suffisant. Pour notre échantillon de  $80 h^{-1}\text{Mpc}$ , la masse entrante et sortante de ce volume (de l’instant initial à l’instant présent) représente déjà 16% de la masse totale. Pour une sphère de  $40 h^{-1}\text{Mpc}$ , le flot de matière est encore plus grand : il peut varier entre 30% et 63% de la masse totale suivant le catalogue 8k-mockX considéré. Bien que les effets de marée et de variance cosmique deviennent plus petit sur une échelle de  $80 h^{-1}\text{Mpc}$ , ils continuent d’affecter de manière importante les bords du domaine lagrangien d’un catalogue donné d’une manière non-triviale. Comme nous allons le montrer, ces problèmes sont amplifiés par les distortions en redshift.

Pour être en mesure de comparer des choses comparables, nous avons effectué une reconstruction sur 8k-mock6 en utilisant le domaine lagrangien donné par la simulation ; cette reconstruction est appelée *TrueDom*. Maintenant, nous confrontons les résultats de *TrueDom* pour deux configurations différentes de reconstruction qui tentent de retrouver le domaine lagrangien :

- la reconstruction *NaiveDom* est obtenue en supposant naïvement un volume lagrangien sphérique pour 8k-mock6. Dans ce cas, toute la masse qui est actuellement dans le catalogue 8k-mock6 était uniformément dans une sphère de rayon  $80 h^{-1}\text{Mpc}$ . De manière équivalente, cela signifie qu’il n’y a pas eu de flux de masse significatif à travers les bords comobiles du catalogue dans le passé.
- La reconstruction *PaddedDom* est obtenue en utilisant un remplissage homogène du volume à l’extérieur du catalogue 8k-mock6. Le remplissage est choisi de manière à ce que le maillage final de MAK qui sera reconstruit est un cube inhomogène (comme dans le panneau de droite de la deuxième ligne des figures 5.14 et 5.15). Le cube doit être suffisamment gros pour absorber les fluctuations de densité présentes au niveau des bords du catalogue (typiquement une zone tampon d’épaisseur  $20 h^{-1}\text{Mpc}$  est requise). Avec de vraies données, nous sommes obligés de supposer que le catalogue est complètement représentatif de la totalité de l’Univers, *i.e.* sa densité moyenne effective est égale à  $\Omega_m$ .

La figure 5.14 montre les résultats des reconstructions *TrueDom*, *NaiveDom* and *PaddedDom* appliquées à 8k-mock6 en l’absence de distortions en redshifts. La figure 5.15 donne les résultats pour les mêmes types de reconstruction mais quand on les applique à des catalogues de redshifts. La table 5.5 résume les valeurs des moments de  $P(v_{r,\text{sim}}, v_{r,\text{rec}})$  dans différents cas. Nous allons maintenant d’abord

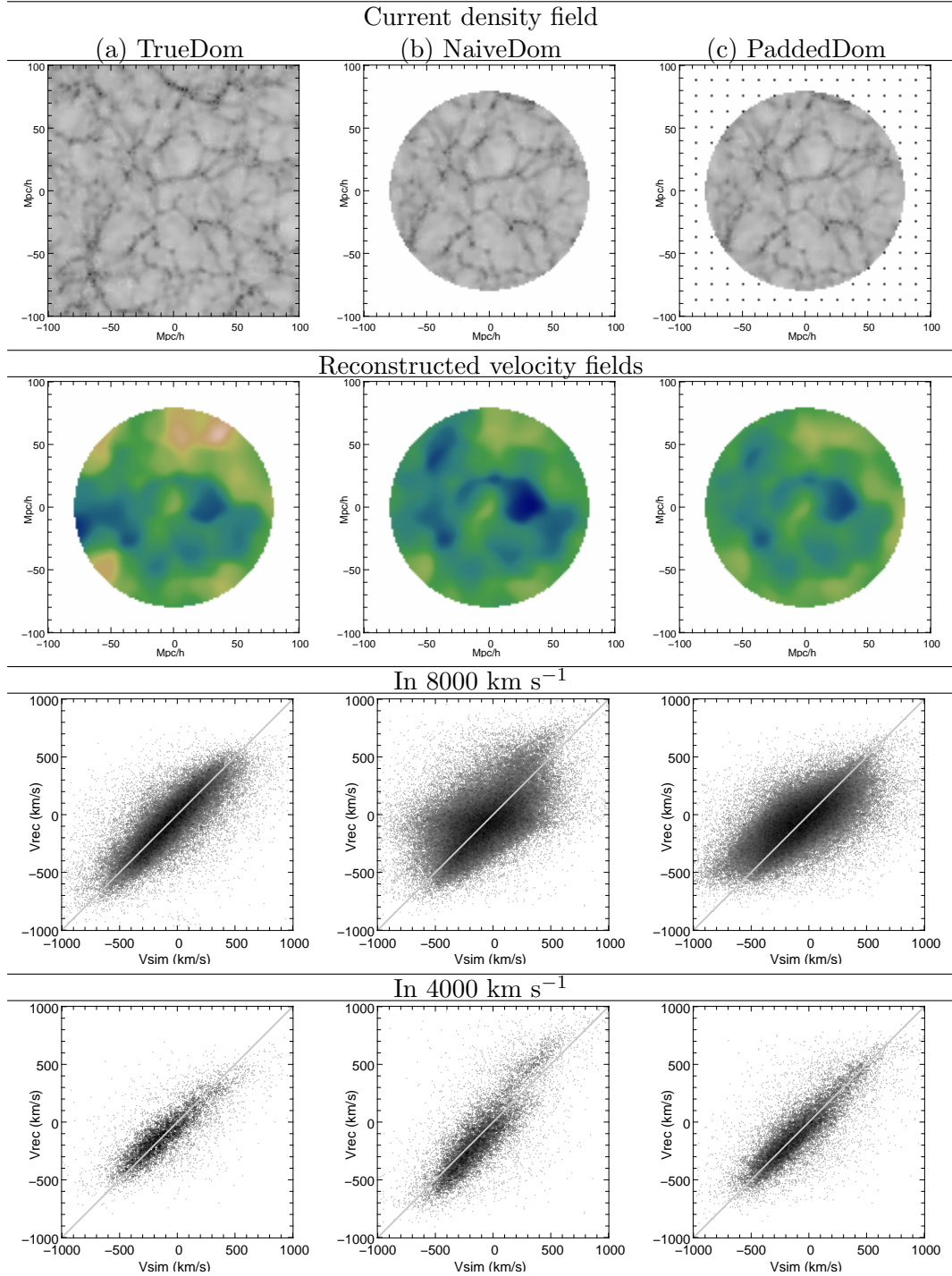


FIG. 5.14: *Domaine lagrangien / sans distortions de redshift* – Cette figure résume les résultats obtenus sur des reconstructions qui ont une information limitée sur le domaine lagrangien. La colonne de gauche illustre la reconstruction *Truedom*, la colonne du milieu celle appelée *NaiveDom*, et celle de droite la reconstruction *PaddedDom*. Pour des raisons d’occupation d’espace, le champ de vitesse original donné par la simulation n’est pas rappelé ici mais peut être trouvé dans la figure 5.8 (page 81). La *ligne du haut* illustre les trois schémas utilisés pour gérer les effets du bord du champ de densité : dans la colonne de gauche, on retient les informations sur les champs de marée à grande échelle, dans la colonne du milieu on coupe le catalogue à l’aide d’une sphère et on fait une reconstruction sur celle-ci, dans la colonne de droite on complète le catalogue coupé avec des particules distribuées de manière homogène sur une grille. La *deuxième ligne* donne le champ de vitesse reconstruit dans chaque cas, lissé avec une fenêtre gaussienne de  $5 h^{-1} \text{ Mpc}$  comme d’habitude. Le code de couleur est le même que pour les autres figures, c’est-à-dire le bleu profond correspond à  $-500 \text{ km s}^{-1}$  et le blanc à  $+500 \text{ km s}^{-1}$ . La *troisième ligne* compare les vitesses reconstruites et simulées individuelles (sans lissage) pour les objets du catalogue 8k-mock6. La *quatrième ligne* montre la même comparaison mais pour les objets se situant dans la région à moins de  $4000 \text{ km s}^{-1}$  de l’observateur du catalogue 8k-mock6.

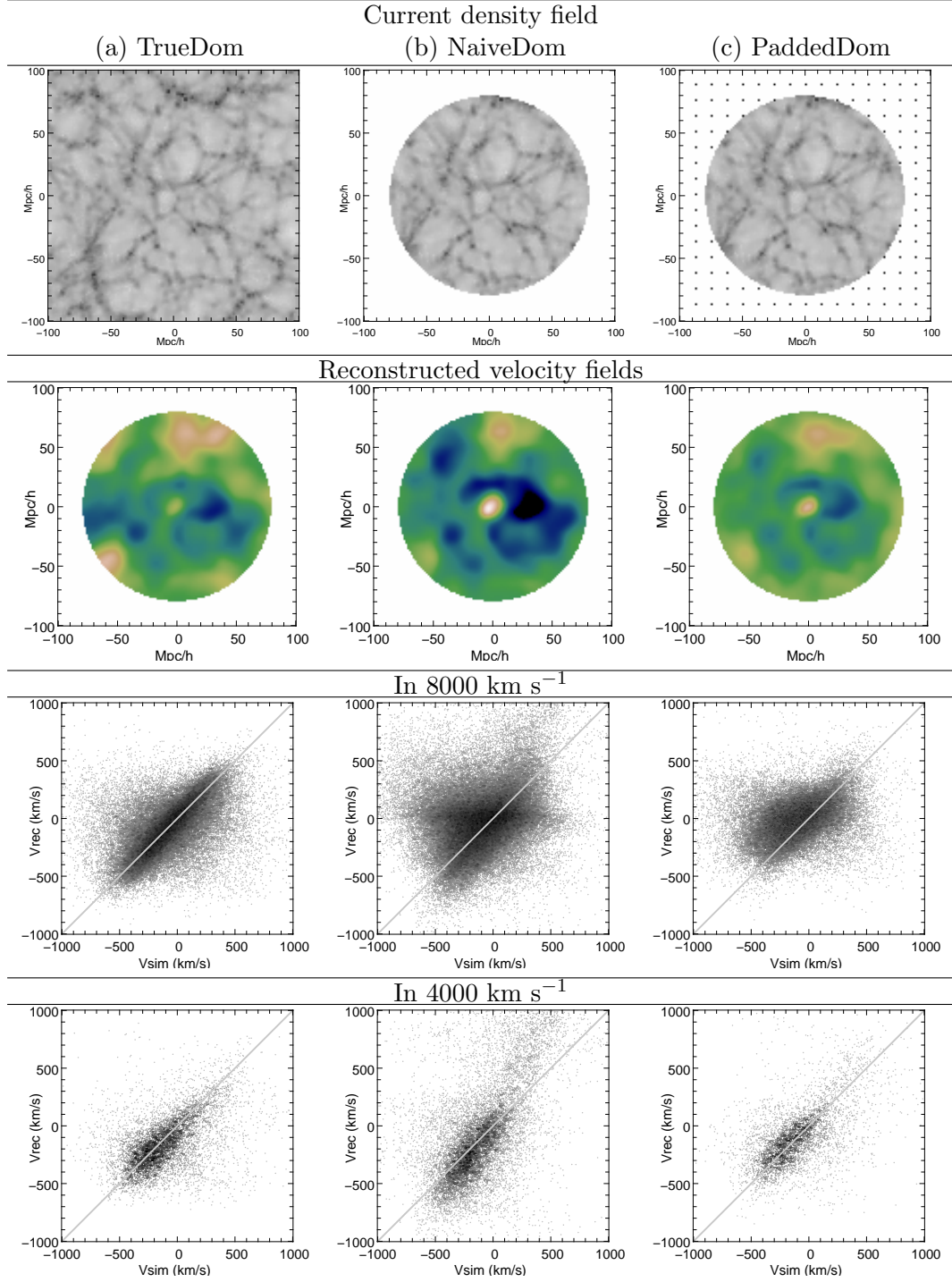


FIG. 5.15: *Domain lagrangien / avec distortions de redshift* – Même chose que pour la figure 5.14, mais pour des catalogues virtuels incluant les distortions de redshift.

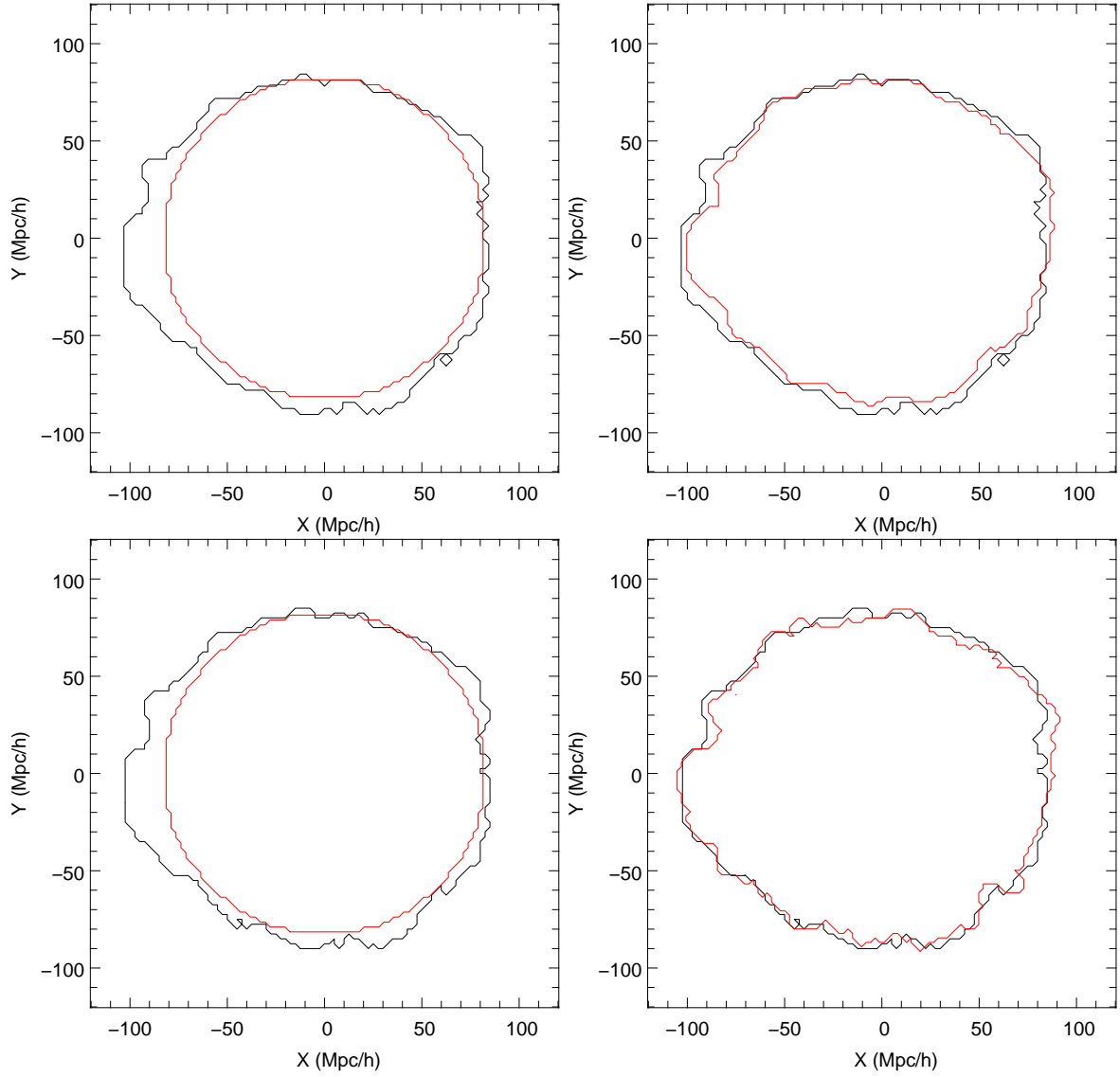


FIG. 5.16: *Volume lagrangien d'un catalogue* – Contour de la projection suivant l'axe Z du volume lagrangien de 8k-mock6 tel que donné par la simulation. On donne pour comparaison (en rouge/gris) le volume NaiveDom correspondant (gauche) et PaddedDom (droite).



confronter les résultats des reconstructions en espace réel et ensuite celles en espace des redshifts.

La reconstruction *TrueDom* ne donne pas de biais significatif à  $80 h^{-1}\text{Mpc}$ . Cependant, à  $40 h^{-1}\text{Mpc}$ , les effets de variance cosmique introduisent une erreur systématique notable dans la direction des plus grands  $\Omega_m$ , ce qui sera discuté dans la section 5.4.3. Comparé à *TrueDom*, *NaiveDom* donne de bons résultats dans l'ensemble bien que la région bleue centrale dans *TrueDom* se transforme en une région bleu sombre dans *NaiveDom*, ce qui suggérerait que le champ de vitesse est biaisé. Cette analyse est confirmée en regardant le diagramme de dispersion du bas. La mesure de  $\Omega_m$  (Table 5.5) est sous-estimée d'à peu près 26% même dans la région centrale du catalogue qui devrait normalement être moins affectée par les effets de bord. D'un autre côté, *PaddedDom* ne semble pas avoir un tel problème au milieu de 8k-mock6, c'est-à-dire dans la région 4k-mock6. A la fois le diagramme de dispersion du bas et la mesure de  $\Omega_m$  (sous-estimation de 8%) confirment que les vitesses reconstruites sont quasiment sans biais dans la région centrale. Comme on s'y attendait, les vitesses au voisinage des bords sont totalement fausses pour les deux méthodes.

A présent, nous coupons les catalogues en espace redshift. Les biais dues à la distortion redshift affectent plus fortement la distribution des vitesses des objets au niveau du bord du catalogue : le catalogue contient en moyenne plus d'objets tombant sur lui que d'objets s'en éloignant. Dans quelques cas, on peut même trouver des objets artificiellement séparés du volume principal du catalogue (ils ont l'air "déconnectés"). L'hypothèse de convexité est alors définitivement perdue pour ces objets. Ce problème amplifie les effets de bord. Le cas des reconstructions *TrueDom* a été discuté dans la section 5.3. Comme précédemment, les vitesses propres dans *NaiveDom* et dans *PaddedDom* sont largement décorréliées dans le volume complet de 8k-mock6 (Fig. 5.15). Cependant, les vitesses propres reconstruites par *NaiveDom* semblent plus fortement sur-estimées qu'en utilisant celles de *PaddedDom*, comme cela est montré par la table 5.5. Pour *NaiveDom*, le nuage de point est entâché d'un alignement horizontal dans la Fig. 5.15, panneau en milieu-bas, ce qui est la signature d'un effet de bord très fort. Ce faux alignement était déjà présent, bien que beaucoup moins apparent, dans le cas des reconstructions en espace réel. D'un autre côté, *PaddedDom* ne présente pas cette caractéristique mais simplement une grosse dispersion. Nous avons vérifié que les objets appartenant à l'alignement horizontal proviennent essentiellement du bord à  $80 h^{-1}\text{Mpc}$ , contrairement aux vitesses reconstruites par *PaddedDom* qui sont plus ou moins distribuées uniformément et essentiellement non corrélées aux vitesses simulées.<sup>9</sup> Cela signifie que *PaddedDom* semble être au moins meilleur dans la gestion des effets de bord que *NaiveDom* dans le sens que les erreurs sont distribuées d'une manière plus aléatoire et moins systématique. Bien que bas de façon impressionnante dans les deux dernière lignes de la table 5.5, le coefficient de corrélation  $r$  est en fait altéré par les longues queues de la PDF que l'on voit dans les dispersions de la figure 5.15. Concernant  $\Omega_m$ , *NaiveDom* semble moins robuste à produire une estimation non biaisée que *PaddedDom*. En effet, en regardant la table 5.5, on peut remarquer que l'intervalle délimité par  $s_{\text{med}}$ ,  $s_{\text{min}}$  et  $s_{\text{max}}$  ne contient "quasiment" pas  $\Omega_m = 0.30$  pour *NaiveDom*/Real space/ $40 h^{-1}\text{Mpc}$ , et ne le contient pas du tout pour *NaiveDom*/Redshift space. Au contraire,  $\Omega_m = 0.30$  est toujours sélectionné par les trois paramètres  $s$  en utilisant une reconstruction *PaddedDom*. Dans le reste de ce travail, partout où cela est nécessaire, nous utiliserons donc la reconstruction *PaddedDom*.

### 5.4.3 La variance cosmique

Nous supposons généralement que les catalogues de galaxies donnent une représentation non biaisée de l'Univers dans son entier, mais bien sûr nous n'avons aucune garantie que cette hypothèse est

<sup>9</sup>Ce comportement est attendu d'un point de vue algorithmique. Les objets proches du bord ne peuvent acquérir tous les déplacements à cause de la "pression"/compétition des objets à l'intérieur de la sphère. Ce problème est amplifié en espace des redshifts car généralement ces objets proviennent de l'extérieur de la sphère et sont sélectionnés, leur vitesse de chute sur le catalogue étant élevée. Avec *NaiveDom*, ils ne peuvent s'échapper du volume lagrangien supposé sphérique ce qui amène à ce que leur vitesse vaille zéro. D'un autre côté, *PaddedDom* est beaucoup moins strict sur le bord, ce qui laisse la liberté à la reconstruction MAK d'avoir une vitesse non-nulle même pour les objets aux bords du catalogue.

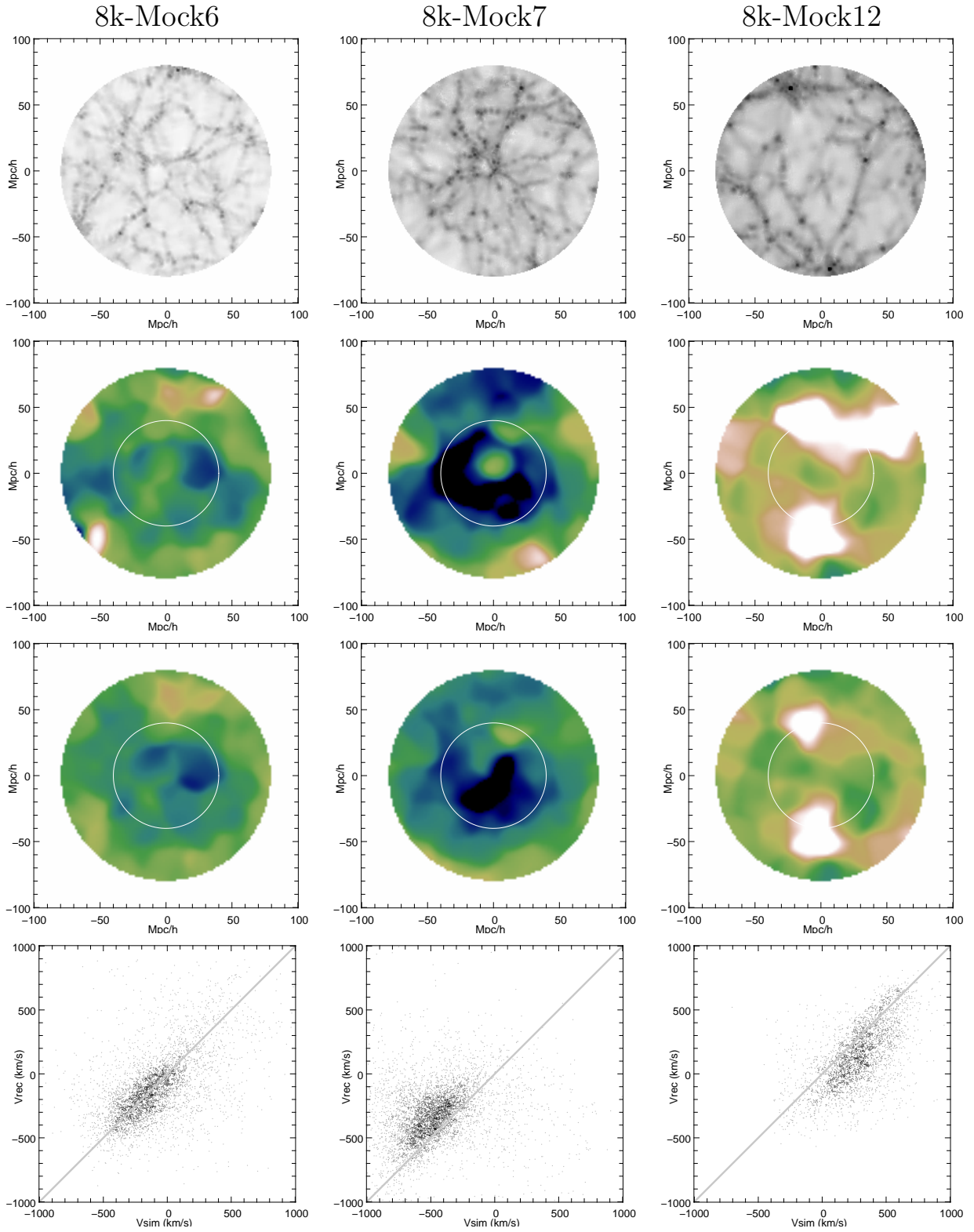


FIG. 5.17: *Variance cosmique* – Cette figure donne une comparaison visuelle des trois catalogues virtuels utilisés pour étudier les effets de variance cosmique. Panneaux du haut : champ de densité des catalogues virtuels considérés, lissés adaptativement. Dans chaque cas, nous avons représenté la tranche infiniment fine centrale qui contient l’observateur. Deuxième ligne : champ de vitesse simulé, après lissage avec une fenêtre gaussienne de  $5 h^{-1}\text{Mpc}$  de rayon. Le cercle blanc donne la limite du volume de  $40 h^{-1}\text{Mpc}$  de rayon. Troisième ligne : la même chose que la deuxième ligne mais pour le champ de vitesse reconstruit. Quatrième ligne : comparaison entre les vitesses propres reconstruites et simulées.

TAB. 5.5: *Volume lagrangien* – Erreur résiduelle après la correction. La description de certaines des colonnes est donnée dans la légende de la table 5.1. “Radius” donne le rayon de la sphère sur laquelle est effectuée la comparaison vitesse-vitesse. “Reconstruction type” indique le type de domaine lagrangien utilisé pour la reconsturction et si celle-ci inclut des effets de distortion en redshift. Les détails de la signification de chacun des noms sont donnés dans la section 5.4.2.

Reconstruction type	Radius ( $h^{-1}\text{Mpc}$ )	Velocities			$\Omega_m$	$\Omega_m$	$\Omega_m$	$\Omega_m$	$\Omega_m$	$\Omega_m$
		$s$	$r$	$\sigma$	( $s$ )	( $\mathcal{L}_{\min}$ )	( $\mathcal{L}_{\max}$ )	$1.5\sigma, s_{\text{med}}$	$1.5\sigma, s_{\min}$	$1.5\sigma, s_{\max}$
TrueDom / Real space	80	0.91	0.77	0.66	0.35	0.28	0.31	0.27	0.233	0.32
	40	0.80	0.76	0.65	0.45	0.28	0.38	0.35	0.28	0.43
NaiveDom / Real space	80	0.87	0.52	0.92	0.38	0.20	0.28	0.42	0.20	0.87
	40	1.11	0.77	0.73	0.25	0.20	0.24	0.244	0.19	0.31
PaddedDom / Real space	80	0.73	0.65	0.77	0.53	0.36	0.48	0.45	0.27	0.75
	40	0.91	0.77	0.64	0.35	0.28	0.34	0.32	0.26	0.38
NaiveDom / Redshift space	40	1.49	0.51	1.31	0.11	0.15	0.26	0.20	0.12	0.37
PaddedDom / Redshift space	40	0.93	0.53	0.94	0.36	0.18	0.34	0.38	0.20	0.79

correcte. Donc, le résultat d’une reconstruction MAK peut être affecté par des inhomogénéités au-delà de l’échelle du catalogue. Par exemple, notre galaxie peut siéger dans une région particulièrement extrême (sur-dense ou sous-dense), ce qui provoquerait des vitesses propres non-usuelles. Cet effet, connu sous le nom de *variance cosmique*, peut être étudié en utilisant nos trois catalogues virtuels de base : 4k-mock6, 4k-mock7, 4k-mock12 (section 5.1). L’effet de variance cosmique est ici amplifié par le caractère fini du volume échantillonné. Dans notre cas, le volume est suffisamment petit pour avoir une vitesse moyenne non nulle sur la ligne de visée. Sur une échelle de  $40 h^{-1}\text{Mpc}$ , cet effet peut modifier substantiellement la mesure de  $\Omega_m$  (écrit  $\Omega_{m,\text{mes}}$  dans cette section) en coupant la distribution  $P(v_{\text{rec}}^s, v_{\text{sim}})$  à un endroit inadéquat.

### Une étude de cas : le collapse sphérique

Nous renvoyons le lecteur à la section 3.5.2 pour la présentation du modèle de l’effondrement sphérique et les équations que nous utilisons pour le résoudre. Nous allons ici nous focaliser sur une partie des résultats que l’on peut en déduire. Pour cela nous avons considéré le problème suivant : nous voulons connaître le facteur de croissance effectif  $\tilde{\beta}$  reliant le champ de vitesse à l’amplitude du champ de déplacement dans le cas de l’effondrement sphérique. Nous cherchons donc à trouver une relation (dans les notations de 3.5.2) :

$$R_0 \dot{a} b'(a) = \tilde{\beta} \frac{\dot{a}}{a} R_0 (1 - b(a)) \quad (5.23)$$

où on a écrit ici une relation linéaire entre le déplacement d’une couche de rayon initial comobile  $R_0$  (membre de droite) et la vitesse de cette couche (membre de gauche). Nous pouvons ensuite utiliser la relation entre  $\Omega_{\text{mes}} = \tilde{\beta}^{9/5}$  pour étudier l’influence de l’environnement sur la mesure du  $\Omega_m$  de l’Univers. Cette comparaison est donnée figure 5.18. Nous y voyons que l’influence de l’environnement est d’autant plus forte que  $\Omega_\Lambda$  est faible. On peut tenter une estimation de l’erreur que l’on va commettre pour un Univers avec  $\sigma_8 = 1$ . Dans ce cas on trouve, avec un spectre de puissance Bardeen et al. (1986), que  $\sigma_{40h^{-1}\text{Mpc}} = 0.19$ , ce qui correspond à une erreur sur  $\Omega_m$  entre 5% et 8% d’après le modèle de l’effondrement sphérique.



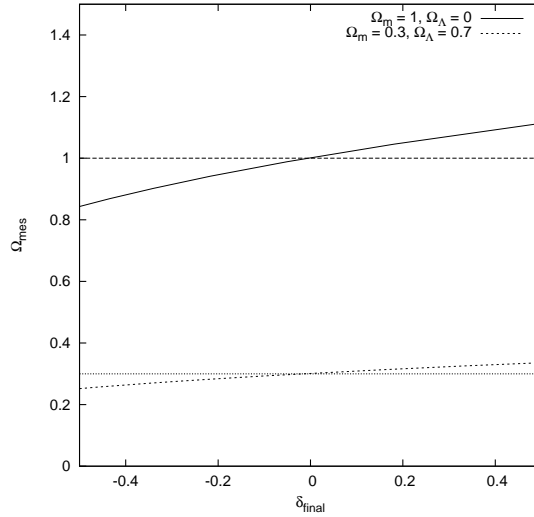


FIG. 5.18: Evolution de  $\Omega_{m,mes}$  en fonction de l'environnement – Deux cas sont ici représentés :  $\Omega_m = 1, \Omega_\Lambda = 0$  (ligne solide) et  $\Omega_m = 0.3, \Omega_\Lambda = 0.7$  (ligne tiret fin). Pour donner une référence de comparaison, il a été aussi représenté les lignes correspondant à  $\Omega_{m,mes} = 0.3$  (pointillé fin) et  $\Omega_{m,mes} = 1.0$  (tiret large). L'axe des abscisses donne l'amplitude de la fluctuation à l'instant présent, comme cela serait le cas pour un observateur.

### Retour sur la reconstruction

Les résultats de la reconstruction sur ces trois catalogues virtuels sont donnés sur la figure 5.17. Dans la table 5.6, nous donnons, pour chaque catalogue virtuel, le meilleur résultat disponible (et donc mettant en avant l'effet de choisir un catalogue virtuel précis) et les résultats qu'on obtiendrait par l'observation de cette portion d'Univers. Les effets de détection du domaine lagrangien, de distortion redshift et d'effets d'incomplétude sont ajoutés à ces catalogues virtuels. Les problèmes d'affectation de masse et de l'obscurcissement du plan galactique sont laissés de côté pour la clarté de l'exposé. Leur impact sur les vitesses devrait très probablement rester le même que ce que nous avons montré dans les sections correspondantes, *i.e.* un biais pour le premier et une augmentation de la dispersion pour le second. Seuls les cas avec les effets observationnels mentionnés ci-dessus sont représentés sur la figure 5.17.

Une inspection visuelle des diagrammes du bas de la figure 5.17 montre que l'effet de volume fini va probablement rendre la mesure  $\Omega_{m,mes}$  sensible au  $\Omega_m$  local ( $\Omega_{eff}$  dans la table) comme on l'a déjà vu dans le cas du collapse sphérique. Cette hypothèse est supportée par l'estimation de  $s$  et  $\Omega_m$  pour des reconstructions *TrueDom* donné dans la table 5.6. De plus, les expériences ci-dessus conduites avec le modèle du collapse sphérique montrent que  $\Omega_{m,mes}$  est en effet une moyenne pondérée entre  $\Omega_{eff}$  et  $\Omega_m$ .

Plus spécifiquement, les vitesses reconstruites dans *4k-mock7* (incluant les effets observationnels) donnent apparemment le  $\Omega_m$  de la simulation mais ils ont une large dispersion ce qui rend l'estimation de la pente douteuse. En effet, si on fait la même reconstruction mais sans les effets observationnels, on obtient  $\Omega_{m,mes} = 0.40$ , ce qui est la moyenne exacte entre le  $\Omega_{m,simu} = 0.30$  de la simulation et  $\Omega_{eff} = 0.50$ .<sup>10</sup> La dispersion mentionnée ci-dessus est par ailleurs attendue pour ce catalogue virtuel : le champ de vitesse n'est pas bien reconstruit près de l'observateur dans ce cas (panneau du milieu) car les flots cosmiques sont plus importants que d'habitude ( $\sim 1000 \text{ km s}^{-1}$ ) et les non-linéarités sont plus fortes. Par conséquent la convexité du problème est perdue pour une région étendue autour de l'observateur quand la reconstruction est faite en espace des redshifts. (voir section 5.3). Une caractéristique

<sup>10</sup>Le collapse sphérique prédirait ici plutôt  $\Omega_{m,mes} = 0.35$  pour la même configuration.

TAB. 5.6: *Variance cosmique* – Résumé des mesures conduites sur les trois catalogues virtuels. La reconstruction est soit effectuée sur le catalogue de base sans effet observationnel, mise à part la variance cosmique (noté *Original*), ou sur le même catalogue mais en ajoutant des effets de distortion de redshift, d’incomplétude et pour lequel le domaine lagrangien est déterminé avec la méthode *PaddedDom* (noté *Full*). La description de chacune des colonnes est donné dans la table 5.1 (page 74).

Catalogue	Reconstr.	Champ de vitesse			$\Omega_m$	$\Omega_m$	$\Omega_m$	$\Omega_m$	$\Omega_m$	$\Omega_m$
		$s$	$r$	$\sigma$	( $s$ )	( $\mathcal{L}_{\min}$ )	( $\mathcal{L}_{\max}$ )	$1.5\sigma, s_{\text{med}}$	$1.5\sigma, s_{\min}$	$1.5\sigma, s_{\max}$
4k-mock6 ( $\Omega_{\text{eff}} = 0.35$ )	Original	0.80	0.76	0.65	0.313	0.28	0.38	0.35	0.28	0.43
	Full	0.94	0.50	0.96	0.35	0.13	0.31	0.31	0.16	0.70
4k-mock7 ( $\Omega_{\text{eff}} = 0.5$ )	Original	0.70	0.67	0.76	0.57	0.39	0.47	0.40	0.33	0.48
	Full	0.88	0.11	1.33	0.43	0.41	1.62	0.30	0.09	1.29
4k-mock12 ( $\Omega_{\text{eff}} = 0.19$ )	Original	1.12	0.81	0.66	0.24	0.235	0.27	0.24	0.22	0.26
	Full	1.08	0.58	1.11	0.24	0.29	0.62	0.15	0.08	0.31

particulièrement importante qui n’est pas reconstruite est celle de la “bulle” en expansion au centre qui disparaît dans le champ de vitesse reconstruit. La taille de la région affectée est d’environ  $20 h^{-1}\text{Mpc}$  autour de l’observateur dans 4k-mock7, ce qui limite donc le nombre d’objets aux bonnes vitesses reconstruites tout en restant observables.

Au contraire, les vitesses dans 4k-mock12 sont reconstruites avec une meilleure corrélation, comme cela est montré par la table 5.6, mais la mesure de  $\Omega_m$  est fortement transformée en direction de  $\Omega_{\text{eff}}$ . Ces deux “caractéristiques” sont principalement dues à l’énorme vide central. Premièrement, la reconstruction MAK et l’approximation de Zel’dovich marchent mieux dans les régions de basse densité qui ne comprennent pas de croisements de coquilles. De plus, le centrage autour d’un vide entraîne l’inhibition des distortions de blueshift car les galaxies sont principalement en train de s’éloigner de l’observateur, ce qui rend convexe le problème de reconstruction dans l’équation (5.21). Deuxièmement, la région de basse densité affecte beaucoup la distribution statistique des vitesses, ce qui dans ce cas amène à mesurer un  $\Omega_{m,\text{mes}}$  pondéré plus fortement en direction du  $\Omega_{\text{eff}}$  de 4k-mock12.<sup>11</sup> Ceci nous amène à un  $\Omega_{m,\text{mes}}$  qui est plus proche de  $\Omega_{m,\text{eff}}$  dans 4k-mock12 que de la densité moyenne de matière dans la simulation complète. Le caractère fini du volume produit aussi un décalage apparent entre les vitesses reconstruites et celles mesurées. On s’y attend car en faisant une analyse statistique sur un catalogue de volume fini, on doit introduire un effet de biais de sélection. Nous avons en effet vérifié que l’ensemble de points  $\{(v_{r,i}, \psi_{r,i}^s)\}$ , obtenu à travers une reconstruction MAK appliquée à 4k-mock12, est un sous-ensemble de points du nuage correspondant à une reconstruction sur 8k-mock12. En regardant notre catalogue virtuel “standard” 4k-mock6, on peut remarquer que la distribution de vitesses simulées est généralement plus symétrique par rapport à la vitesse nulle que pour les deux autres catalogues virtuels, et sans biais visuel quand on compare les vitesses reconstruites à celles simulées. Cela vient renforcer notre hypothèse initiale liant  $\Omega_{m,\text{mes}}$  à  $(\Omega_{m,\text{simu}}, \Omega_{\text{eff}})$  et la distribution asymétrique de vitesses. Potentiellement, on pourrait calculer le vrai  $\Omega_m$  de l’Univers (ou ici de la simulation) à partir des vitesses mesurées de n’importe quel catalogue en prédisant comment l’asymétrie de la distribution de vitesse est liée au contraste de densité local. Cependant, la solution la plus simple et la plus robuste consiste encore à étendre la profondeur des catalogues existant jusqu’à atteindre un volume tel que la distribution des vitesses soit distribuée une distribution normale.

D’un point de vue prédictif, la comparaison visuelle des champs de vitesse à l’intérieur des cercles blancs montre que, si l’on connaît  $\Omega_m$ , nous pouvons reconstruire des champs de vitesse plausibles pour les trois catalogues virtuels. En dehors des cercles blancs, la comparaison visuelle des champs de

<sup>11</sup>Le collapse sphérique prédirait un  $\Omega_{m,\text{mes}} = 0.26$  mesuré et ceci est en bon accord avec la valeur mesuré quand aucun effets observationnel n’est injecté dans le catalogue virtuel.

vitesse aux champs obtenus par simulation montre une absence de corrélation quasi complète, comme nous en avons discuté dans la précédente section. Il faut remarquer que le champ de vitesse tend doucement vers zéro (couleur verte) sur le bord de tous les catalogues virtuels : c’est un effet attendu du remplissage homogène qui tend à lisser toute fluctuation au bord (champ de vitesse et de densité).

## 5.5 Erreurs sur les mesures de distance

Le dernier effet observationnel que nous traiterons dans ce chapitre est lié aux erreurs de mesure des vitesses observées. Comme nous allons le voir ces erreurs sont suffisamment importantes pour fausser la mesure de  $\Omega_m$ .

### 5.5.1 A-t-on réellement besoin d’une analyse de vraisemblance ?

Tous les effets décrits précédemment sont présents dans un catalogue de redshift. Bien qu’on s’attende à ce que la plupart des biais observationnels soient indépendants, certains pourraient être corrélés et donner des erreurs systématiques bien pires. Nous présentons sur la figure 5.19 la détérioration progressive de la comparaison vitesse-vitesse pour le catalogue *4k-mock6* basé sur une reconstruction faite sur *8k-mock6*. Les effets sont empilés de gauche à droite. Les mesures de  $\Omega_m$  pour la méthode  $1.5\sigma$  sont indiqués en-dessous de chaque panneau. La conclusion évidente est que ces mesures sont progressivement affectées mais qu’aucune erreur corrélée supplémentaire ne semble être introduite quand on mélange les effets. Au contraire il semble que, par chance, les biais semblent se contrebalancer pour donner à la fin un résultat non biaisé (avant dernier panneau). Si on effectue une reconstruction *TrueDom/redshift* au lieu de *TrueDom/real*, on tend à faire décroître  $\Omega_m$  comme on l’a vu précédemment. Au contraire, injecter les effets d’incomplétude remonte la mesure vers des  $\Omega_m$  plus élevés comme on l’a remarqué dans la section 5.2.3. La méthode  $1.5\sigma$  semble donner le bon  $\Omega_m$  dans tous les cas, ce qui signifie qu’on devrait être capable de l’utiliser sur des catalogues de galaxies pourvu qu’on ait une précision suffisante sur les mesures de vitesses propres. Cependant, si on regarde le dernier panneau (en bas à droite) de la figure 5.19, il semble que si on prend en compte des erreurs aléatoires sur les mesures de vitesse (ici nous avons introduit une erreur optimiste de 8% de la distance à l’objet, correspondant à une erreur sur le module de distance de  $\sigma_\mu = 0.17$ ), l’estimation de la pente semble beaucoup plus difficile. Dans ce cas, le  $\Omega_m$  mesuré est fortement biaisé. On s’y attend car la méthode  $1.5\sigma$  s’appuie principalement sur la partie centrale du nuage, qui est dans ce cas la partie qui est la plus affectée par les erreurs aléatoires. Cela amène une circularisation du contour  $1.5\sigma$  et donc une estimation totalement fautive de la pente. D’un autre côté, si on s’intéresse à la structure globale du nuage, on s’aperçoit que la pente attendue est encore cachée dans les données, mais que si on souhaite la trouver il faut prendre en compte les queues de la distribution. Ce dernier test montre les limites d’une comparaison directe des vitesses reconstruites aux vitesses observées. Il serait possible de retrouver la distribution originale de la dispersion en déconvoluant le bruit. Cependant, il semble que cela soit une opération difficile et nous préférons d’abord essayer une approche en maximum de vraisemblance. Son principal avantage est de travailler directement sur les distances, ce qui rend les erreurs dans les mesures plus gérable. Cette analyse a été présentée dans la section 4.3.1. Nous allons donc nous concentrer à présent sur les résultats d’une telle approche dans des cas réel (contrairement à la section 4.3.3) dans la section suivante.

### 5.5.2 Résultat

Les résultats de la mesure de  $\Omega_m$  en utilisant un estimateur “maximum de vraisemblance” sont présentées dans les tables en utilisant le libellé  $\mathcal{L}$ . Excepté dans le cas où nous considérons les erreurs observationnelles, nous utilisons la version simplifiée de  $\mathcal{L}$  en prenant  $\sigma_{0,i} = 0$ . Alors qu’il serait naturel de marginaliser par rapport à tous les paramètres (ce qui inclue  $B_v$ ,  $\sigma_v$ ,  $\gamma_*$ ), nous avons rapidement remarqué que cela entraînait des mesures biaisées d’une manière inacceptable et à une augmentation inutile de l’espace des paramètres. De plus, les résultats dépendent assez fortement de  $\sigma_v$  et  $\gamma_*$ , tout

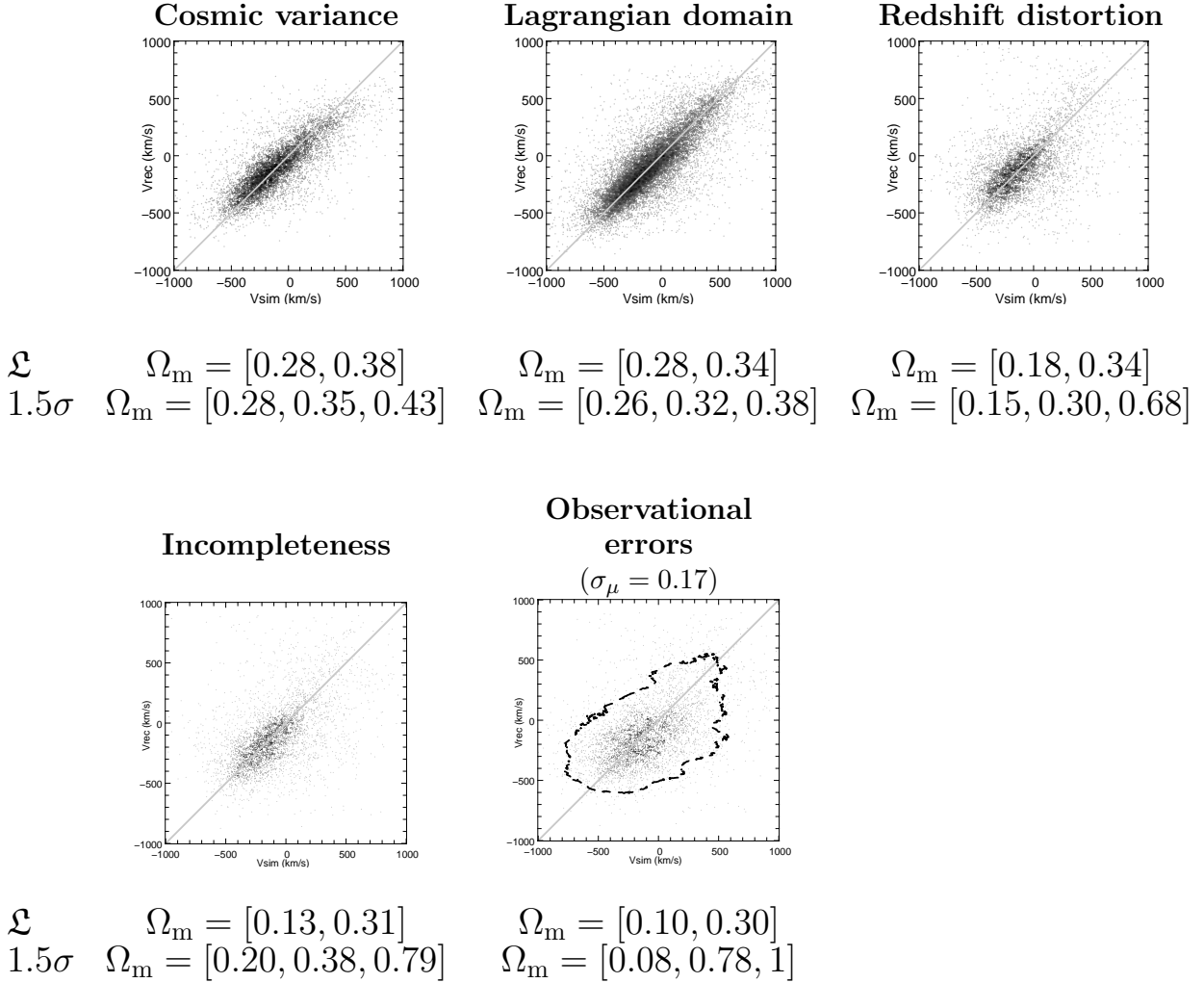


FIG. 5.19: This figure gives the evolution of the scatter distribution and of the measurement of  $\Omega_m$  using it while more and more observational effects are added to 8k-mock6 catalogue. All measurements of  $\Omega_m$  are given between brackets and are sorted as follows. For measurements obtained through the likelihood analysis, labelled by  $\mathcal{L}$ , the first number corresponds to  $\mathcal{L}_{\min}$  and the second to  $\mathcal{L}_{\max}$ . For measurements obtained using the  $1.5\sigma$  method, the first number corresponds to  $s_{\min}$ , then  $s_{\text{med}}$  and finally  $s_{\max}$ . The last (lower right) panel uses the full likelihood function of Eq. (4.29). All others use a restricted likelihood analysis with  $\sigma_0/e = 0$ , which is nearly equivalent to using Eq.(4.28) for each  $(v_{r,i}, \psi_{r,i})$  pair. The  $1.5\sigma$  isocontour has been plotted with a thick dashed line in the last panel.

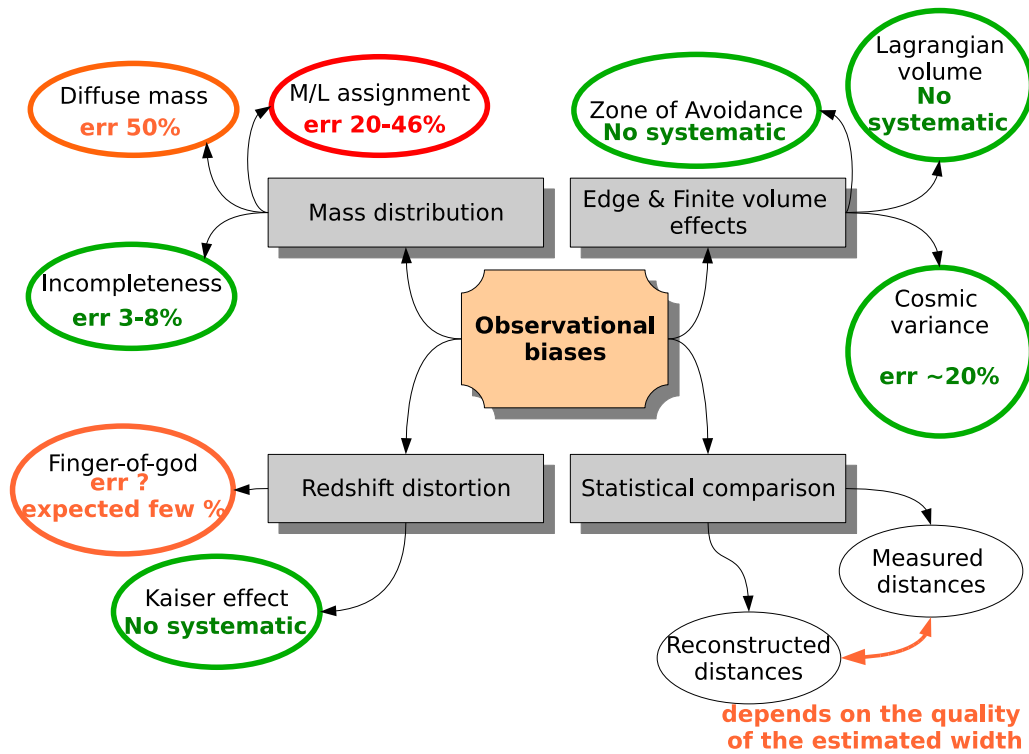
spécialement quand le bruit de reconstruction est assez élevé comme dans les reconstruction redshift (voir 4.3.3). Nous proposons donc, pour discuter les valeurs obtenues à prendre  $\gamma_* = 0$ ,  $B_v = 90 \text{ km s}^{-1}$  et à choisir deux valeurs pour  $\sigma_v$ . Dans un premier cas, nous utilisons la théorie linéaire pour prédire la dispersion moyenne des vitesses des amas dans l’Univers (cf. section C), ce qui nous amène à prendre  $\sigma_v = 326 \text{ km s}^{-1}$  (le  $\Omega_m$  mesuré de cette manière est libellée  $\mathcal{L}_{\text{max}}$ ). Dans un deuxième cas,  $\sigma_v = +\infty$  est choisi pour vérifier l’influence du prior sur la distribution de vitesse, la mesure correspondante est libellée avec  $\mathcal{L}_{\text{min}}$ .

En regardant toutes les tables de ce chapitre, nous remarquons que la différence entre les deux  $\Omega_m$  mesurés suit principalement l’intervalle défini par  $s_{\text{min}}$  et  $s_{\text{max}}$ . Nous nous attendions à un tel comportement ( $\sigma_v$  contrôle plus ou moins le biais statistique de la fonction de vraisemblance) mais pas qu’il suivrait aussi clairement l’autre méthode. Plus la dispersion est importante, plus la mesure devient imprécise comme on s’y attend. Il faut cependant remarquer qu’en moyenne la mesure  $\mathcal{L}_{\text{max}}$  souffre moins de biais systématique que  $\mathcal{L}_{\text{min}}$ . Ce comportement est confirmé par les tests que nous avons fait à la section 4.3.3. Le  $\Omega_m$  qui est apparemment bien estimé dans le panneau en bas à droite de la figure 5.19 a été calculé en utilisant l’analyse complète en fonction de vraisemblance ( $\sigma_{0,i} \neq 0$ ). En fait, si on compare à la méthode  $1.5\sigma$  pour laquelle la pente mesurée n’est pas définie,  $\mathcal{L}_{\text{min}}$  et  $\mathcal{L}_{\text{max}}$  ont en fait la *même valeur* que dans le cas où nous n’avions *pas* introduit *d’erreurs observationnelles*. On ne peut, bien sûr, faire confiance à la correction basée sur l’hypothèse d’une distribution de vitesse dans le cas de 4k-mock7 et 4k-mock12. Comme on peut le remarquer sur la figure 5.17, la distribution de vitesse n’est pas du tout gaussienne dans ces cas-là. Cela invalide notre modélisation de la section 4.3.1. En regardant la table, on remarque que, bien que les mesures sur les reconstructions “Original” ne sont pas fortement affectées, nous ne pouvons dire la même chose en utilisant les données des reconstructions “Full”. Dans le premier cas, le bruit est suffisamment bas pour que le prior n’ait pas beaucoup d’importance alors que dans le second une modélisation fausse de la distribution de vitesse nous amène à une erreur énorme sur le  $\Omega_m$  mesuré. Heureusement, la dispersion de points comporte d’autres propriétés qui restaurent les mesures de la table 5.6 entre l’estimation avec la fonction de vraisemblance ( $\sigma_v = +\infty$  afin de supprimer le prior gaussien) et la méthode  $1.5\sigma$ . Pour 4k-mock7 et 4k-mock12, l’estimation de la pente est facilitée par la présence de nombreux traceurs munis d’une grande vitesse, ce qui diminue la possibilité d’une erreur systématique sur la pente.

On est donc amené en définitive à utiliser des catalogues de distance suffisamment profonds pour nous assurer d’une distribution de vitesse proche d’une gaussienne afin de pouvoir appliquer la correction sur la fonction de vraisemblance. Dans ce cas, on peut s’appuyer sur la valeur donnée par  $\mathcal{L}_{\text{max}}$ . Si, au contraire, la distribution de vitesse est non-gaussienne, on doit utiliser  $\mathcal{L}_{\text{min}}$ . Si possible, on doit faire une inspection visuelle du nuage de points de la comparaison vitesse-vitesse afin de vérifier qu’aucun biais statistique majeur n’entache le résultat.

## 5.6 Conclusion sur les systématiques

La carte des effets systématiques assortis des estimations quantitatives des erreurs typiques commises sur  $\Omega_m$  dans le cadre des catalogues virtuels considérés, est présentée ci-après à titre de conclusion.



Cette synthèse apporte une réponse aux questions que nous nous étions posées dans la partie introductive de ce chapitre. Comme prévu, la précision des estimations est relativement limitée par la méconnaissance de la fonctionnelle qui associe une distribution de masse à une distribution de luminosité. Il nous faudra donc étudier ceci de façon plus détaillée dans l'avenir.



# 6 Le champs de vitesse du 2MASS Redshift survey (2MRS)

## Sommaire

<b>6.1</b>	<b>Le catalogue de distance NBG-3k . . . . .</b>	<b>103</b>
6.1.1	Présentation du catalogue . . . . .	103
6.1.2	Traitement du catalogue en vue de la reconstruction . . . . .	104
6.1.3	Le champs de vitesse reconstruit . . . . .	106
<b>6.2</b>	<b>Le catalogue de galaxies avec décalage vers le rouge : 2MASS Redshift Survey . . . . .</b>	<b>107</b>
6.2.1	Présentation du catalogue . . . . .	107
6.2.2	Distributions des galaxies du catalogue . . . . .	110
6.2.3	Préparation du catalogue : phase I . . . . .	110
6.2.4	Préparation du catalogue : phase II . . . . .	119
6.2.5	Comparaison du champ de densité du NBG-3k et du 2MASS . . . . .	124
6.2.6	Champ de vitesse reconstruit . . . . .	124
<b>6.3</b>	<b>La vitesse d'ensemble locale et l'origine du dipôle du CMB . . . . .</b>	<b>125</b>
6.3.1	Mesurer la vitesse d'ensemble : le méthode $\chi^2$ . . . . .	125
6.3.2	Analyse de la vitesse d'ensemble dans une simulation $\Lambda$ CDM . . . . .	130
6.3.3	La vitesse d'ensemble observée par le catalogue NBG-3k . . . . .	133
6.3.4	La vitesse d'ensemble reconstruite à partir du catalogue 2MRS . . . . .	134
<b>6.4</b>	<b>Comparaison aux vitesses observées et paramètres cosmologiques . . . . .</b>	<b>136</b>
6.4.1	Analyse des vitesses simulées et reconstruites d'une simulation $\Lambda$ CDM . . . . .	145
6.4.2	Vitesses observées et reconstruites dans le catalogue 2MRS/NBG-3k . . . . .	145
<b>6.5</b>	<b>Conclusion . . . . .</b>	<b>147</b>

Nous allons ici utiliser les méthodes de reconstruction du champ de vitesse qui ont été développées aux chapitres précédents pour obtenir celui de notre Univers local. Pour cela, nous allons utiliser les données fournies par deux catalogues de galaxies : un catalogue de distance (et de décalages vers le rouge) appelé NBG-3k (Nearby Galaxy catalog – 3000 km s<sup>-1</sup>deep, section 6.1), et un catalogue beaucoup plus profond de décalages vers le rouge appelé 2MASS Redshift Survey (2MRS, section 6.2). Nous utiliserons les résultats données par la reconstruction MAK sur ce dernier catalogue pour obtenir des informations sur l'origine du mouvement du Groupe Local par rapport au fond diffus cosmologique ainsi que sur celui du superamas local (section 6.3.4). Enfin, nous ferons une comparaison du champ de vitesse reconstruit à celui observé par le catalogue NBG-3k à la section 6.4. Nous validerons la procédure utilisée dans ces deux dernières parties à l'aide d'un catalogue virtuel, construit à l'aide de trois contraintes environnementales pour ressembler à notre Univers local.

## 6.1 Le catalogue de distance NBG-3k

### 6.1.1 Présentation du catalogue

Le catalogue NBG-3k est un agrandissement de 40% du “Nearby Galaxy Catalog” (Tully, 1987; Tully et al., 2008). Il inclut 3300 galaxies à l'intérieur d'un volume de 30  $h^{-1}$ Mpc. Il a une bonne complétude quand on le compare à d'autres catalogues comme le 2MASS. Il est constitué de deux composantes. La première est un catalogue de redshifts (décalage vers le rouge) de galaxies :



- les galaxies sont regroupées en groupes ou filaments. Si une galaxie semble isolée dans le catalogue, elle est laissée en dehors de tout groupe.
- des galaxies virtuelles sont introduites à basse latitude galactique afin d’éviter des effets de bord dans la ZOA (section 5.4.1 et 6.2.4).
- une correction en luminosité est appliquée pour tenir compte de l’extinction “naturelle” des objets de plus faibles luminosité à grande distance, causée par la sélection des objets selon leur magnitude apparente lors de la construction du catalogue.

La deuxième composante, portant sur un sous-ensemble de la première, est un catalogue de distances de galaxies. Parmi les 3300 galaxies, on a pu affecter un module de distance à  $\sim 1800$  galaxies (section 4.3.1 et 2.4.2) à l’intérieur du volume de  $30 h^{-1}\text{Mpc}$  de rayon. Pour cela quatre indicateurs de distance ont été utilisés : les étoiles Céphéides (Freedman et al., 2001b), la “Tip of the red giant branch” (Karachentsev et al., 2003; Lee et al., 1993), les variations de brillance de surface (Tonry & Schneider, 1988; Tonry et al., 2001) et la relation de Tully-Fisher (Tully & Fisher, 1977; Tully & Pierce, 2000). Parmi ces galaxies, 600 ont une distance mesurée avec un des trois premiers indicateurs qui ont une précision assez supérieure à la relation de Tully-Fisher. On utilisera par la suite la version groupée du catalogue, avec donc des distances moyennées sur ces groupes. En effet, les reconstruction MAK n’étant pas à même de retrouver les régions avec croisement de coquilles, il est plus avantageux de travailler directement sur les groupes. Le catalogue présente alors 1234 groupes, qui peuvent ne contenir qu’une galaxie, et parmi ceux-ci 633 ont eu leur distances mesurées. Un graphique représentant la position des groupes dans ce catalogue est donnée sur la figure 6.1. On donne aussi une représentation du champ de densité de masse de ce catalogue, en utilisant l’algorithme de filtrage adaptatif (Annexe E), sur cette même figure. Le champ de masse a été produit en plaçant les objets à leur position redshift et en leur affectant une masse liée à la luminosité des galaxies elliptiques et spirales :

$$M_i = L_{i,\text{spirale}} \times 100 + L_{i,\text{elliptique}} \times 300 , \quad (6.1)$$

avec  $i$  désignant un objet du catalogue. L’amplitude absolue du rapport  $M/L$  n’a pas d’importance comme on l’a déjà discuté précédemment, par contre le rapport entre le  $M/L$  des elliptiques et celui des spirales en a. Ici, on affecte donc trois fois plus de masse aux elliptiques qu’aux spirales. Ce choix est motivé par des applications du théorème du viriel à des galaxies (Mohayaee & Tully, 2005; Tully, 2005).

### 6.1.2 Traitement du catalogue en vue de la reconstruction

On applique les algorithmes qui ont été vus tout au long des chapitres précédents. Comme le catalogue est fourni avec une correction de luminosité, on ne se préoccupera pas de l’amplification de la luminosité dans un premier temps. On utilise l’affectation en masse telle qu’elle a été précisée par l’équation (6.1). On a décidé ici de faire deux reconstructions : une dans le référentiel du groupe local,<sup>1</sup> et une autre dans le référentiel du CMB. Les vitesses correspondant au décalage vers le rouge des galaxies ne sont pas les mêmes et donc la reconstruction peut donner deux résultats différents. Les deux référentiels ont tous les deux des avantages et des inconvénients :

- le référentiel du groupe local ne donne pas de galaxies décalées vers le bleu et qui ne peuvent être reconstruites. Il est plus naturel pour l’étude des catalogues car il ne fait pas l’hypothèse qu’à grande échelle le référentiel des grandes structures et du CMB convergent.
- le référentiel du CMB présente l’avantage que nous ne sommes pas un point particulier dans l’espace. Le champ de vitesse reconstruit doit avoir une moyenne tendant vers zéro à grande échelle, ce qui doit correspondre au résultat de la reconstruction quand on se place dans le régime linéaire (par exemple Juszkiewicz et al., 1990). Donc c’est un référentiel, sous l’hypothèse qu’il est égal au référentiel de repos des grandes structures, qui est asymptotiquement, sur les grands volumes, le bon pour la reconstruction. Néanmoins, pour des reconstructions de volume fini on

<sup>1</sup>Le groupe local est défini grossièrement, dans Karachentsev et al. (2002), à partir des galaxies incluses dans le volume défini par la surface de redshift héliocentrique nul.

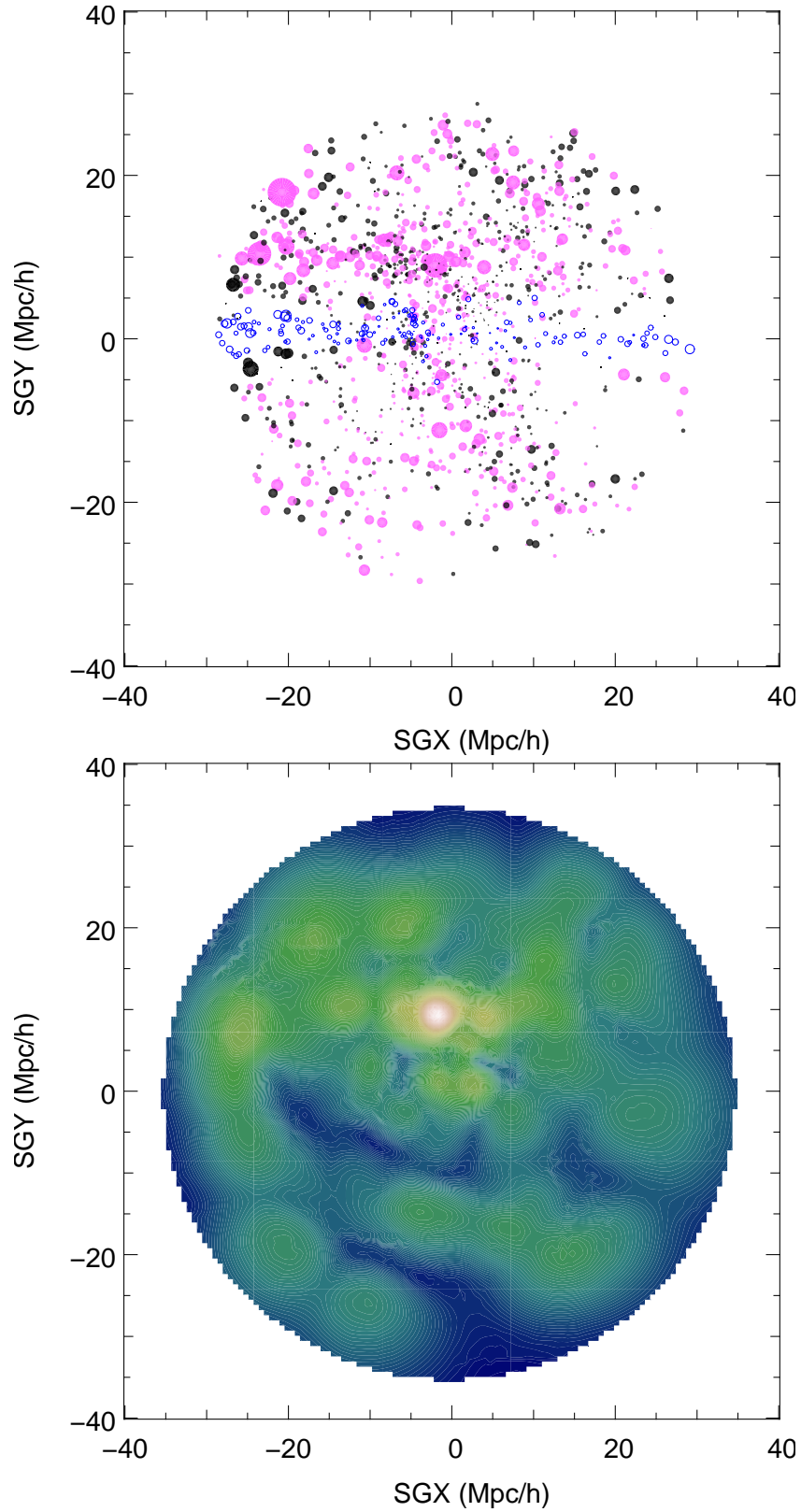


FIG. 6.1: *Distribution de masse dans le catalogue NBG-3k* – Panneau du haut : projection selon l’axe Z supergalactique de la position des groupes du NBG-3k. Les disques noirs indiquent la position de vrais groupes, mais dont seul la distance redshift est connue. Les disques magenta indiquent les groupes dont la distance lumineuse a été mesurée. Les cercles bleus non remplis indiquent la position des groupes virtuels introduits pour remplir la ZOA. Dans tous les cas la taille du cercle est proportionnelle à la racine cubique de la masse. Panneau du bas : coupe du champ de densité à  $SGZ=0$  (coupe infiniment fine du champ, l’échelle de lissage est de l’ordre de quelques  $h^{-1}\text{Mpc}$ ). Le blanc indique des régions de forte densité. Le bleu indique des régions de sous-densités.

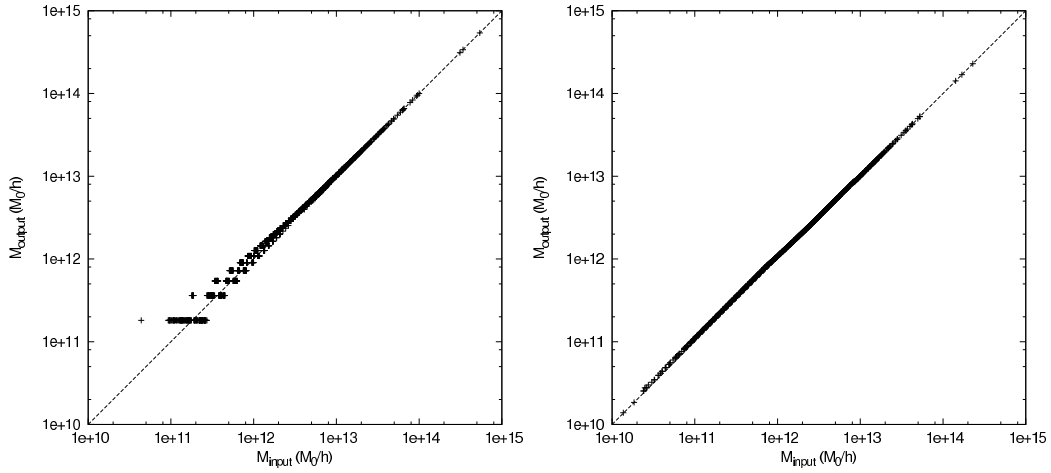


FIG. 6.2: On mesure un biais de 2% pour la reconstruction  $64^3$  particules, et un biais inférieur à 2 pour mille pour la reconstruction  $140^3$ .

peut avoir un doute sur la signification exacte des vitesses reconstruites (c'est-à-dire par rapport à quel référentiel elles sont exprimées).

Des tests plus exhaustifs seront faits à ce propos dans la section 6.3.1. On suppose pour la reconstruction que  $\Omega_m = 0.258$  (Dunkley et al., 2008) pour corriger des distortions dues aux vitesses propres. Dans le maillage de MAK qui est construit à partir du catalogue, on omet explicitement la contribution de la galaxie d'Andromède (M31) et de la Voie Lactée. En effet, ces deux galaxies ont un décalage vers le rouge négatif et ne peuvent donc avoir leur position corrigée avec la formule (5.21). On a choisi un maillage de résolution  $N_{\text{MAK}} = 140^3$  particules pour échantillonner un volume de  $100^3 h^{-3} \text{ Mpc}^3$  après remplissage, ainsi que  $N_{\text{MAK}} = 64^3$ . Un choix aussi élevé de nombre de particules est nécessaire afin d'utiliser tous les groupes disponibles du catalogue et de réduire le biais dans la construction du maillage de MAK donné par l'algorithme 3.3.2. Dans le cas choisi ici, le biais numérique mesuré est inférieur à 1%. On a représenté la relation entre la “masse” de l'objet dans le maillage de MAK et sa masse d'origine sur la figure 6.2 pour les cas  $64^3$  et  $140^3$ . On s'attend en fait uniquement à de très faibles différences entre les deux reconstructions. Nous avons vérifié que le champ de vitesse reconstruit ainsi que la vitesse d'ensemble du catalogue NBG-3k sont les même dans les deux cas. Nous pouvons donc utiliser des résolutions inférieures avec une certaine dose de sérénité.

### 6.1.3 Le champs de vitesse reconstruit

Pour représenter les champs de vitesse tridimensionnels, on utilise généralement le système de coordonnées SuperGalactique. Il s'agit d'un système de coordonnées basé sur le plan approximatif du Superamas de la Vierge. Le plan supergalactique passe par le Soleil et le milieu de l'amas de la Vierge. L'avantage de ce système est que plusieurs autres superamas “proches” se trouvent aussi dans ce plan (tels que l'amas du Centaure, la concentration Shapley, ..., cf. section 2.1). On a représenté sur la figure 6.3 les vitesses propres des groupes du NBG-3k obtenues en appliquant l'algorithme de reconstruction à un maillage  $64^3$ . On constate que les vitesses sont toutes orientées en direction du Grand Attracteur (Hydra et Centaurus) situé à  $\text{SGX} \sim 3000 \text{ km s}^{-1}$  et  $\text{SGY} \sim 1000 \text{ km s}^{-1}$ .

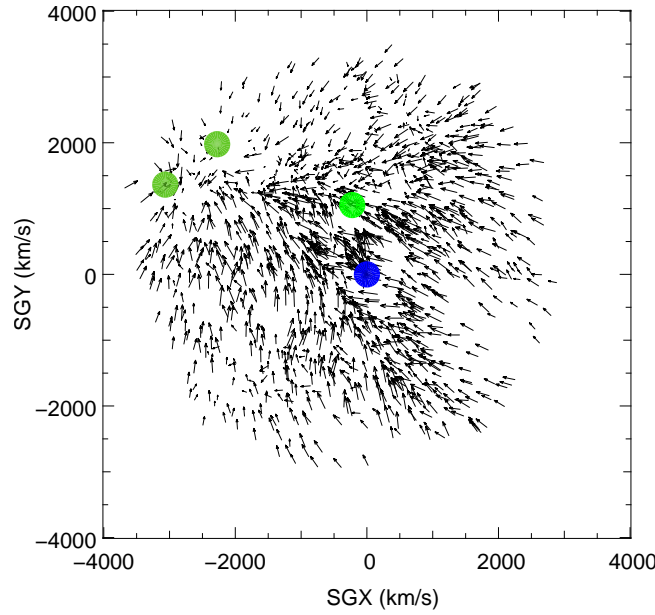


FIG. 6.3: Reconstruction  $64^3$  sur le catalogue NBG-3k, mis dans le référentiel du CMB et projeté selon l'axe supergalactique SGZ. On a représenté la position de la Voie Lactée (en bleu), Virgo (vert au centre), Hydra et Centaurus (vert à gauche de la figure).

## 6.2 Le catalogue de galaxies avec décalage vers le rouge : 2MASS Redshift Survey

### 6.2.1 Présentation du catalogue

Le catalogue 2MASS Redshift Survey (2MRS) est composé de 24761 galaxies sélectionnées parmi les galaxies du 2MASS Extended Source Catalog (2MASS-XSC), telles que leur magnitude apparente en bande  $K_S$  soit  $K_S < m_{\text{limite}} = 11.25$ . Traduite en luminosité solaire cette limite donne

$$S_{\text{limite}} = 10^{0.4(M_{\odot,K} - m_{\text{limit}} + 25)} = 3.82 \cdot 10^6 \text{Mpc}^{-2} L_{\odot}, \quad (6.2)$$

avec  $M_{\odot,K} = 3.39$  (Johnson, 1966). C'est le plus grand relevé en redshifts spectroscopiques couvrant tout le ciel disponible à ce jour. Les galaxies de l'hémisphère céleste nord ont été observées avec le télescope de 1.5 mètres de l'observatoire Fred Lawrence Whipple (FWLO), le télescope d'Arecibo de 305 mètres et le télescope de 100 mètres de Green Bank. Dans l'hémisphère sud, la plupart des galaxies à haute latitude galactique ( $\sim 6000$  galaxies) ont été observées dans le cadre du relevé 6dFGS et les galaxies de basse latitude ont été observées au CTIO (par L. Macri and J. P. Huchra). Une projection Aitoff, en coordonnée galactique, du catalogue 2MRS complet est donné figure 6.4. On a aussi représenté figure 6.5 la distribution du nombre de galaxies en fonction de la distance dans le catalogue 2MASS Redshift Survey. Cette distribution pique à environ  $80 h^{-1} \text{Mpc}$ . A cette distance, les effets d'incomplétude deviennent dominants et on commence à perdre un grand nombre de galaxies. La complétude du catalogue 2MASS-XSC est excellente pour les magnitudes, relativement faibles, des galaxies sélectionnées pour le 2MRS. Pour des hautes latitudes galactiques, typiquement moins de 2% du ciel est masqué par des étoiles. L'uniformité de la photométrie est meilleur que 4% sur le ciel, en incluant les pôles célestes (e.g. Jarrett et al., 2000, 2003). La collaboration du 2MRS a classifié les galaxies du 2MASS dont la luminosité est plus élevée que  $K = 11.25$  : il y a 40% de type précoce (par opposition à 30% dans un échantillon sélectionné en bande B comme le CfA1). Enfin, et le plus important, le catalogue est très complet, y compris aux basses latitudes galactiques  $b \gtrsim 5$  degrés en dehors du bulbe et  $b \gtrsim 10$  dans la direction du bulbe.

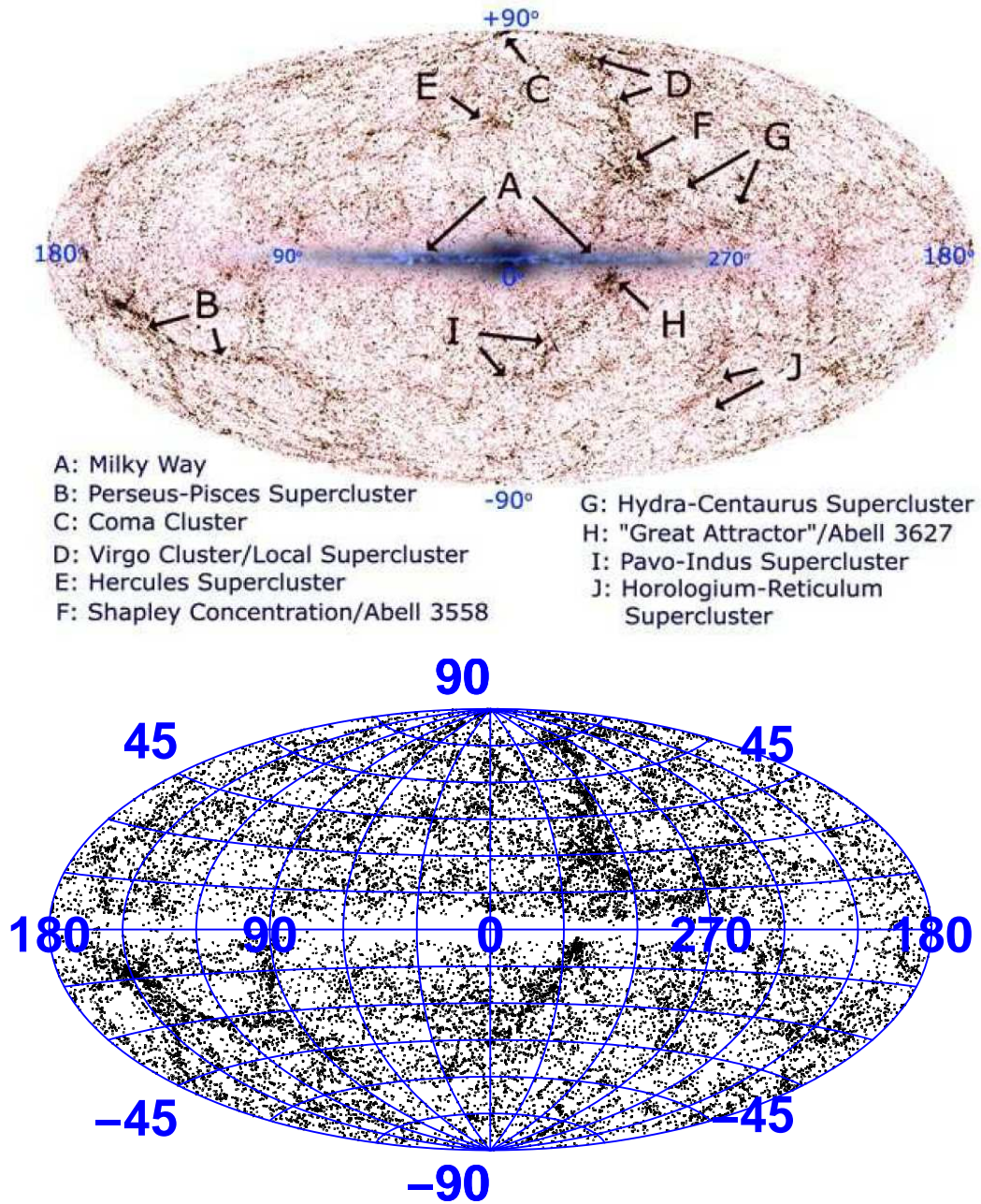


FIG. 6.4: 2MASS Redshift Survey – Projection Aitoff des galaxies du relevé. Les lignes bleues “horizontales”, correspondantes à des isolatitude galactiques, sont espacées de 15 degrés. Les lignes bleues “verticales”, correspondantes à des isolongitudes galactiques, sont espacées de 30 degrés. Seules les galaxies dont le redshift a été mesuré ont été représentées ici. Le bulbe galactique est clairement visible à  $l = 0$ . L'image du panneau du haut a été prise de Huchra et al. (2005).

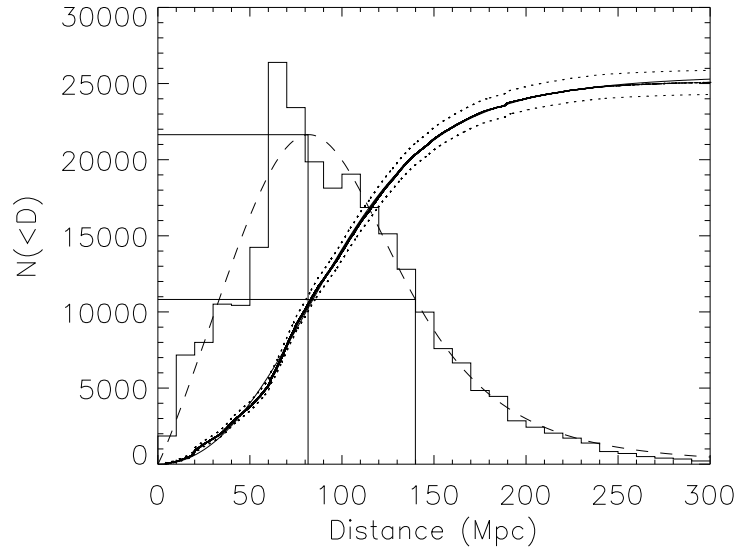


FIG. 6.5: *2MASS Redshift Survey* – Figure prise de Crook et al. (2007). Les petits points indiquent le nombre cumulé de galaxies à l'intérieur d'une sphère de rayon  $D$ . La courbe solide fine a été obtenue par un meilleur ajustement des paramètres de l'équation (6.10). Les courbes en pointillés donnent les barres d'erreur à  $5\sigma$  d'une statistique de Poisson. On a un léger écart aux données à la fois à petite et grande distance. On a représenté en unités arbitraires la dérivée de cette fonction par rapport au rayon, ce qui donne une idée de la densité radiale de galaxies. La fonction issue de l'ajustement admet un maximum à une distance d'environ  $80 h^{-1}\text{Mpc}$ , et elle rejoint une valeur moitié de la valeur du maximum à environ  $140 h^{-1}\text{Mpc}$ .

### 6.2.2 Distributions des galaxies du catalogue

La fonction de luminosité des galaxies du 2MRS a été estimée par Huchra et al. (2005). Ils ont obtenu :

$$\Phi(L) = \frac{\Phi^*}{L^*} \left( \frac{L}{L^*} \right)^\alpha e^{-L/L^*} \quad (6.3)$$

avec

$$\alpha = -1.02, L^* = 5.33 \cdot 10^{10} h^{-2} L_\odot, \Phi^* = 1.06 \cdot 10^{-2} h^3 \text{ Mpc}^{-1}. \quad (6.4)$$

On a reproduit figure 6.6 la distribution de luminosité qu'ils ont estimé. A partir de cette fonction il est possible d'obtenir une prédiction sur le nombre de galaxies à une distance donnée  $R$  présente dans ce catalogue :

$$dN(R) = \int_{L_{\min}(R)}^{+\infty} \Phi(L) dL \times 4\pi R^2 dR \quad (6.5)$$

$$= 4\pi R^2 dR \int_{s_{\min} R^2}^{+\infty} \Phi(L) dL \quad (6.6)$$

$$= 4\pi R^2 \Phi^* \Gamma(1 + \alpha, S_{\text{limite}} R^2) dR. \quad (6.7)$$

avec

$$\Gamma(a, x) = \int_x^{+\infty} t^{a-1} e^{-t} dt \quad (6.8)$$

Donc le nombre cumulé de galaxies dans une sphère de rayon  $R$  est :

$$N(R) = \int_0^R dN(R) = \int_0^R 4\pi R^2 \Phi^* \Gamma(1 + \alpha, \frac{S_{\text{limite}} R^2}{L^*}) dR \quad (6.9)$$

Un meilleur ajustement est donné dans Crook et al. (2007) pour la distribution en redshift des galaxies :

$$N_{\text{C07}}(R) = N_0 \left( \frac{\nu R}{((\nu R)^b + S^b)^{1/b}} \right) \quad (6.10)$$

avec  $N_0 = 2.57 \cdot 10^4$ ,  $\nu = 0.881$ ,  $S = 106 \text{ Mpc}$ ,  $b = 3.94$ . Sur la figure 6.7, on a représenté les courbes donnant le nombre de galaxie à l'intérieur de la sphère de rayon  $R$  donné par l'axe des abscisses pour la valeur de  $L^*$  donnée par Huchra et al. (2005). Il est important d'avoir exactement la bonne correction d'incomplétude ici afin d'éviter le plus possible des effets systématiques dans le champ de vitesses radiales reconstruites. Nous y reviendrons dans la section 6.4.

### 6.2.3 Préparation du catalogue : phase I

Nous allons à présent préparer le catalogue 2MRS pour à la fois une reconstruction précise et une utilisation plus conviviale pour les comparaisons qui suivront les différentes reconstructions. Nous allons appeler ce catalogue *2MRS\_Rec*.

#### Association des galaxies du NBG-3k à celle du 2MRS

La première étape est de trouver le sous-ensemble de galaxies du catalogue NBG-3k (Tully et al., 2008) qui est inclus dans le catalogue 2MRS. Pour cela, nous avons utilisé une base de donnée fournie par Brent Tully qui donne l'identification entre une galaxie du catalogue NBG-3k et une préversion du catalogue 2MRS. Les noms dans le catalogue 2MRS correspondent à leur coordonnées équatoriales<sup>2</sup> exprimé à l'époque J2000. Malheureusement, la précision des coordonnées donnant les noms des galaxies n'est pas la même dans les deux catalogues. Associer les distances du catalogue de Brent Tully

<sup>2</sup>RA (ascension droite) et DEC (declinaison)



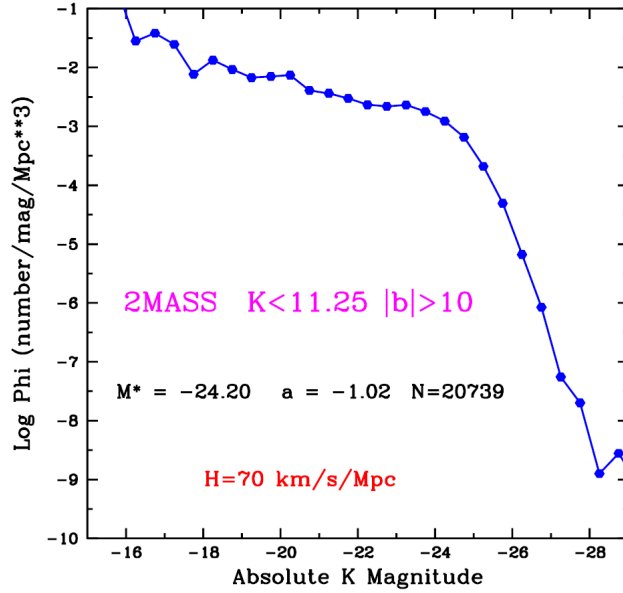


FIG. 6.6: 2MASS Redshift Survey – Distribution en luminosité absolue des galaxies dans le 2MRS. On s’est limité à ajuster la distribution pour des latitudes galactiques  $|b| > 10$  deg. Cette figure est tiré de Huchra et al. (2005).

aux galaxies du 2MRS a été donc un tout petit peu plus délicat que de simplement rechercher un même nom de galaxie dans les deux catalogues. Nous avons donc créé un programme qui tente de trouver des noms de galaxie voisins au sens des coordonnées qui composent le nom. Deux galaxies sont identifiées si “l’écart angulaire” entre les deux noms est de l’ordre de grandeur de l’arrondi. L’objet O1 du catalogue NBG-3k est identifié à l’objet O2 du catalogue 2MRS s’il satisfait à ces deux conditions :

1. il existe une correspondance dans la base de donnée fournie par Brent Tully
2. l’écart angulaire entre l’objet O1 et O2 est minimal.

Après affectation, on a vérifié manuellement que l’écart angulaire entre les deux objets correspondaient bien aux erreurs d’arrondis numériques. Les galaxies appartenant aux deux catalogues ont été représentées figure 6.8 en vert, les galaxies appartenant seulement au catalogue 2MRS ont été représentées en noir. Partant d’un catalogue de distance de 1787 galaxies de l’Univers local, il nous reste, à la fin de l’identification, 1127 galaxies appartenant aux deux catalogues.

### Compression des “Fingers-of-god”

Pour éviter d’être dominé par les régions triplement valuées (voir section 5.3, page 79), il est nécessaire de comprimer les “Fingers-of-god”. Ces distortions, visibles en coordonnées redshift, sont principalement centrées sur les amas de galaxies. Dans ces structures, les vitesses des galaxies sont maxwelliennes et élevées ( $\sim 1000 \text{ km s}^{-1}$ ). Comme ces galaxies sont à peu près toutes à la même distance, et que la distance redshift mélange la vitesse propre des galaxies à leur distance par la relation (2.21), on a la création d’une structure en forme de doigt. Plusieurs algorithmes dérivés du “Friend-Of-Friend” existent (Huchra & Geller, 1982; Tegmark et al., 2004a). Comme cela a été abordé section 3.2, de tels algorithmes regroupent des particules (ou ici des galaxies) résidant dans des régions dont la densité est au-dessus d’un certain seuil. Ici, nous utilisons une variante de cet algorithme qui a été développé par Huchra & Geller (1982). On suppose qu’on se donne une constante  $V_F$  (en  $\text{km s}^{-1}$ ) et  $D_0$  (en Mpc).

1. On choisit un couple de deux galaxies ( $G1, G2$ ) non précédemment traitées. On note respectivement  $z_1$  et  $z_2$  le redshift, exprimé en  $\text{km s}^{-1}$ , de ces deux galaxies. On note aussi,  $D_{1,2}$  leur



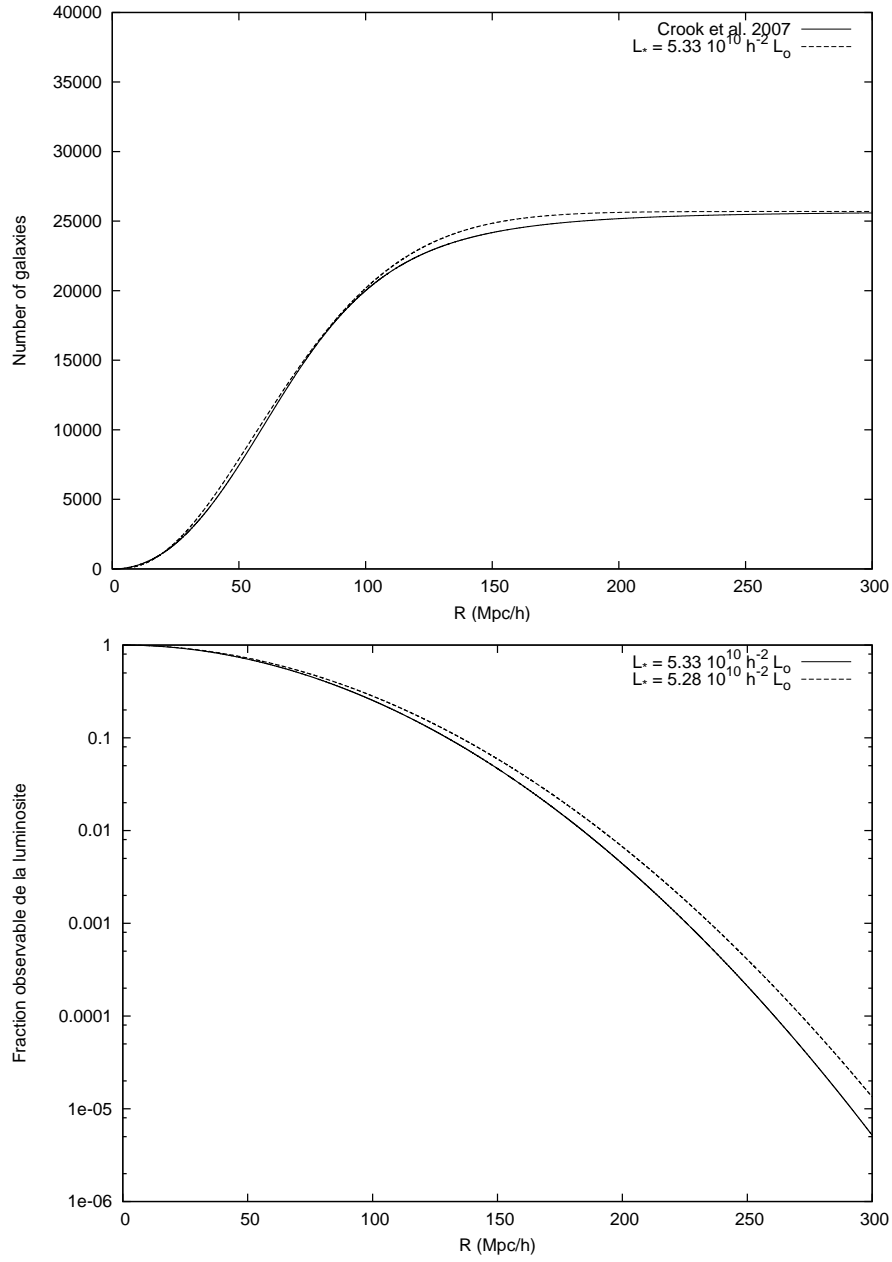


FIG. 6.7: *Comptage de galaxies* – Le panneau de gauche représente le nombre de galaxies à l'intérieur d'un volume de rayon  $R$ . La courbe solide correspond au tracé du meilleur ajustement trouvé par Crook et al. (2007). La courbe en tiret court est la même fonction mais pour  $L_* = 5.33 \cdot 10^{10} h^{-2} L_\odot$ . La distribution corrigée des effets de masque dans le catalogue 2MRS est donnée en pointillés.

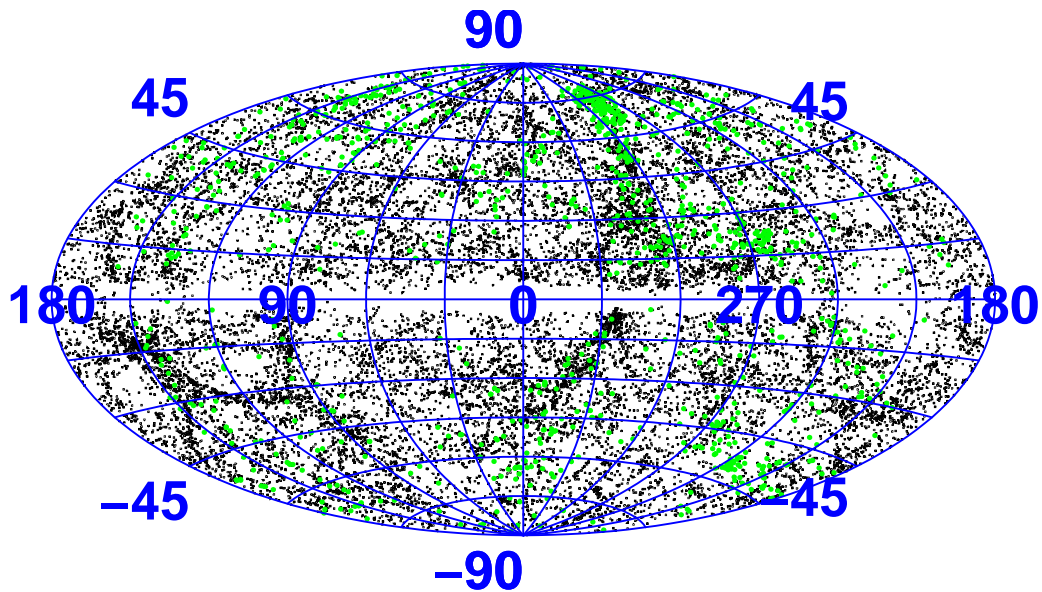


FIG. 6.8: 2MASS Redshift Survey + Catalogue de distance. Les points verts indiquent les galaxies pour lesquelles des mesures de distances sont disponibles.

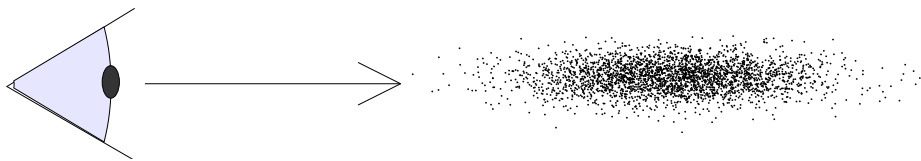


FIG. 6.9: Illustration d'un effet de doigt de Dieu. L'observateur (l'oeil à gauche) observe un amas de galaxies (un point noir représente une galaxie). On a simulé la formation d'une telle structure en coordonnées redshift en supposant que la dispersion de vitesse des galaxies est de  $1000 \text{ km s}^{-1}$  alors que la taille spatiale caractéristique de l'amas est  $2h^{-1} \text{ Mpc}$ .

séparation projetée sur le ciel. Pour calculer  $D_{1,2}$ , on utilise la distance redshift moyenne des deux objets multipliée par le sinus de la séparation angulaire. On a donc, en posant  $\mathbf{u}_1$  et  $\mathbf{u}_2$  les vecteurs unitaires dirigés dans la direction de O1 et O2 :

$$D_{1,2} = 2V_{\text{avg},1,2} \sin \left( \frac{\cos^{-1}(\mathbf{u}_1 \cdot \mathbf{u}_2)}{2} \right), \quad (6.11)$$

avec  $V_{\text{avg},1,2} = |z_1 + z_2|/2$ . Un dessin explicitant ces notations est donné figure 6.10.

2. Si  $|z_1 - z_2| < V_F$  et  $D_{1,2} < D_L(V_{\text{avg},1,2})$  avec

$$D_L(V) = D_0 \left( \frac{\int_{L_{\text{limite}}(V)}^{+\infty} \Phi(L) dL}{\int_{L_{\text{cat},\text{limite}}}^{+\infty} \Phi(L) dL} \right)^{-1/3} = D_0 \left( \frac{\Gamma \left( 1 + \alpha, \frac{L_{\text{limite}}(V)}{L^*} \right)}{\Gamma \left( 1 + \alpha, \frac{L_{\text{cat},\text{limite}}}{L^*} \right)} \right)^{-1/3}, \quad (6.12)$$

avec

$$M_{\text{limite}} = m_{\text{limite}} - 25 - 5 \log_{10} \frac{V}{H_0} \quad (6.13)$$

$$L_{\text{limite}}(V) = 10^{0.4(M_{\odot,K} - M_{\text{limite}}(V))} \quad (6.14)$$

alors  $G1$  et  $G2$  appartiennent au même groupe.

Cet algorithme est asymétrique sur la manière de traiter la distance redshift et la séparation angulaire. On pourrait naïvement penser qu'il faut changer l'échelle de compression en redshift de la même manière que l'échelle de compression angulaire. Néanmoins, Crook et al. (2007) ont montré que la compression des FOG dépend essentiellement du paramètre  $D_0$  dans cet algorithme et très peu de  $V_F$ . On fixera donc  $V_F = 350 \text{ km s}^{-1}$ , paramètre qui donne de bons résultats sur la statistique des groupes dans le CfA (Ramella et al., 1997). Les contours d'isodensité numérique de galaxies sont donc contrôlés par  $D_0$  principalement. On peut estimer la densité moyenne numérique des galaxies qui sont associées à l'aide de l'algorithme de percolation par le calcul suivant :

$$\frac{\delta n}{\bar{n}} = \frac{n_{\text{lié}}}{\bar{n}} - 1 = \frac{3/(4\pi D_L(V_{\text{avg}})^3)}{\int_{L_{12}(V_{\text{avg}})}^{+\infty} \Phi(L) dL} - 1 \quad (6.15)$$

$$= \frac{3}{4\pi D_0^3} \left( \int_{L_{\text{lim}}} \Phi(L) dL \right)^{-1} - 1. \quad (6.16)$$

Nous discuterons dans la suite les résultats sur deux catalogues obtenus par cet algorithme. Un premier catalogue, HDC ("High Dense Catalog"), a été obtenu en choisissant  $D_0$  de telle manière à ce que  $\delta n/\bar{n} = 215$ . Le deuxième, MDC ("Medium Dense Catalog"), a été obtenu pour un  $D_0$  tel que  $\delta n/\bar{n} = 80$ . On obtient respectivement, en utilisant la formule ci-dessus,  $D_0 = 0.4 \text{ Mpc}$ , pour  $\delta n/\bar{n} = 215$ , et  $D_0 = 0.56 \text{ Mpc}$  pour  $\delta n/\bar{n} = 80$ .<sup>3</sup>

Le 2MRS-Rec est donc constitué des groupes et galaxies isolées identifiées. Nous utiliserons pour les séries de reconstructions finales le MDC. Mais entre temps nous ferons quelques comparaisons des propriétés du MDC et du HDC.

### Correction de la mesure de la luminosité totale

Maintenant que les groupes et les modules de distance ont été identifiés, on peut tenter une estimation de la luminosité absolue et totale des groupes (pris au sens large, un groupe peut ne contenir qu'une galaxie) du catalogue 2MRS\_Rec. Pour cela deux cas se présentent :

<sup>3</sup>Il faut remarquer ici que nous ne suivons pas les notations de Crook et al. (2007), où ils avaient opté HDC pour  $\delta n/\bar{n} = 80$  et LDC pour  $\delta n/\bar{n} = 2.28$ .

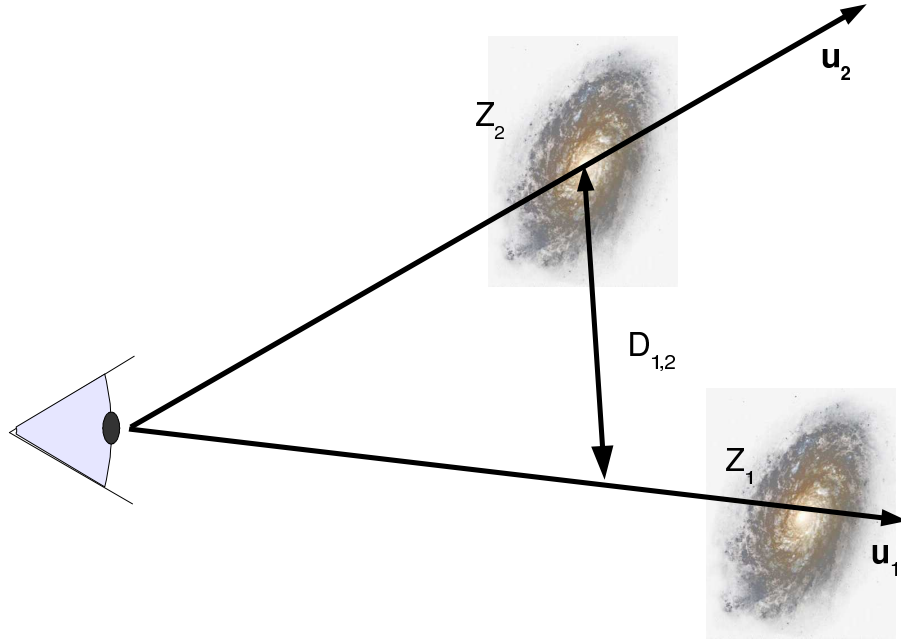


FIG. 6.10: Représentation imagée des paramètres de l'algorithme de compression des finger-of-god

1. Si le groupe comporte une ou plusieurs galaxies dont la distance a été mesurée, on moyenne les modules de distance. Cela donne le module de distance du groupe. On somme les luminosités apparentes des galaxies du groupe. En multipliant par la distance au carré cette luminosité totale, on obtient la luminosité absolue.
2. Si le groupe ne comporte aucune galaxie dont la distance a été mesurée, nous pouvons alors utiliser le redshift moyen du groupe comme un estimateur de sa distance. Contrairement à précédemment, cette opération nécessite une hypothèse sur  $h$ , la constante de Hubble normalisée à  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Cette méthode présente le désavantage d'être excessivement sensible aux vitesses propres pour des objets dans le "voisinage" de l'observateur (typiquement  $\sim 10 h^{-1} \text{ Mpc}$ ). A moins de  $3 h^{-1} \text{ Mpc}$ , toutes les galaxies du 2MRS ont eu leur distance mesurée. On peut donc utiliser celle-ci pour avoir une estimation précise de leur luminosité sans avoir recours au décalage vers le rouge.

Le deuxième point de la liste ci-dessus est vrai dans la dernière version de la base de donnée. Malheureusement, au moment de sa conception (fin 2007) le catalogue MDC n'avait pas accès à toutes les distances. Les galaxies dont la vitesse était inférieure à  $300 \text{ km s}^{-1}$  sont placées à  $3 h^{-1} \text{ Mpc}$  pour éviter de surestimer leur luminosité. Néanmoins tant qu'aucun biais majeur n'est inséré dans 2MRS\_Rec pour notre voisinage (par exemple une surestimation flagrante de la luminosité), aucun problème ne devrait en découler.

Jusqu'à présent, nous avons supposé que la luminosité apparente disponible dans le catalogue 2MASS Redshift survey correspondait à la luminosité apparente *totale* de l'objet. Malheureusement, il n'est pas possible en pratique d'intégrer le flux d'une galaxie pour obtenir cette luminosité, car on pourrait alors confondre la luminosité provenant de plusieurs objets. Dans le catalogue des sources étendues 2MASS-XSC, on se limite à mesurer le flux contenu dans une isophote elliptique pour laquelle le flux moyen intégré sur cette surface de  $20 \text{ mag arcsec}^{-2}$ . On obtient alors la magnitude  $K_{20}$ . Cette luminosité ne donne forcément qu'une limite basse à la luminosité totale de l'objet. Mais il serait souhaitable qu'il ne s'agisse pas d'un estimateur biaisé. En effet, les objets proches ont tendance à avoir des luminosités de surface plus faibles (par exemple pour les galaxies naines) et on pourrait s'inquiéter qu'il existe un biais les concernant. Une des corrections proposées dans le catalogue 2MASS est d'extrapoler cette luminosité observée en utilisant un profil de luminosité connu qui a été ajusté au

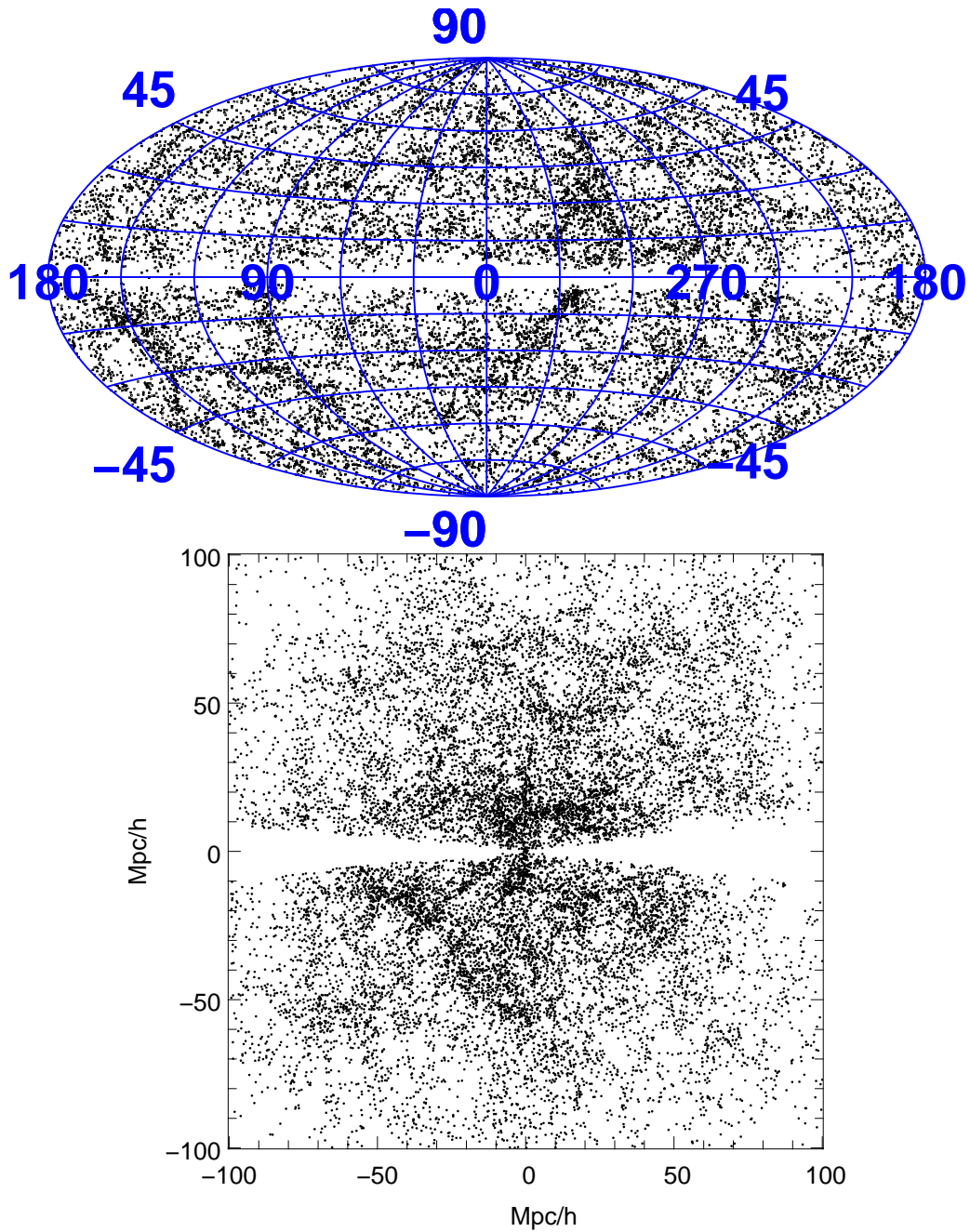


FIG. 6.11: 2MASS Redshift Survey, catalogue de distance, compression des “Fingers-of-god”. Il s’agit du catalogue MDC. Le panneau du haut montre la projection Aitoff du catalogue. On constate que certaines régions ont été légèrement vidées. La compression se voit mieux dans une projection cartésienne (panneau du bas), où le catalogue semble avoir moins de structures pointant vers l’observateur, au centre, que sans compression.

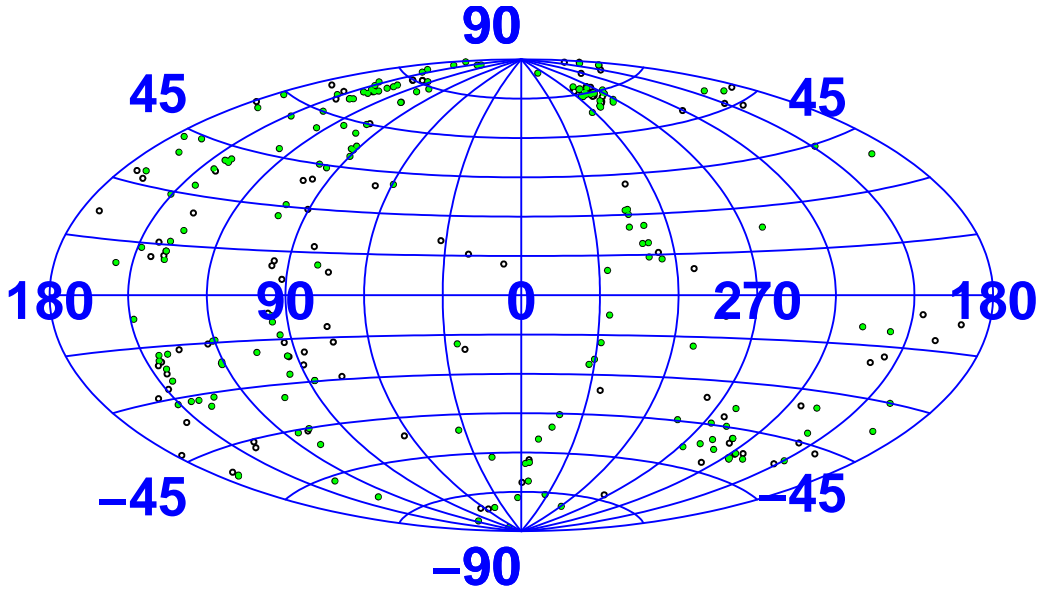


FIG. 6.12: Ce graphique donne la position des galaxies dont le redshift (dans le référentiel du CMB) est inférieur à  $1000 \text{ km s}^{-1}$ . Les galaxies vertes sont celles pour lesquelles la distance n'est pas connue (méthode 1 115). Les points noirs donnent la position des autres galaxies, dont cette fois la distance est connue, dans notre voisinage. Il y a 104 galaxies sans mesures de distance parmi 279 galaxies ( $\sim 38\%$ ).

profil de la galaxie observée. On peut essayer de mesurer le biais en comparant la magnitude apparente  $K_{20}$  à la magnitude estimée  $K_{\text{tot}}$ . Cette comparaison est donnée figure 6.13. Comme on s'y attend aux magnitudes élevées (luminosité faible), la dispersion de  $K_{\text{tot}} - K_{20}$  en fonction de  $M_{\text{tot}}$  (la magnitude absolue totale de la galaxie) augmente avec  $M_{\text{tot}}$  croissant. Cependant, aucun biais global ne semble apparaître sur cet échantillon issu du 2MRS, bien que cela ne signifie pas que pour l'ensemble du 2MASS il n'y ait pas de biais non plus. Nous pouvons donc supposer avec une relative confiance que  $K_{20}$  est un estimateur non biaisé de la luminosité totale de l'objet considéré (à une certaine constante multiplicative qui correspond au décalage vertical du nuage de point dans ce graphique).

### Estimation du rapport masse-luminosité

Comme il a été discuté sur les catalogues virtuels, une bonne estimation de la fonction  $M/L$  est un élément critique afin d'avoir une estimation non-biaisée de  $\Omega_m$ . Nous pouvons nous appuyer d'abord sur les travaux de Lin et al. (2004). Ils ont estimé, en utilisant une relation entre la "température" du gaz de l'amas  $T_X$  et la masse, les variations de la fonction  $\Upsilon_{200} = M_{200}/L_K$  avec  $M_{200}$  pour  $10^{14} M_\odot \lesssim M_{200} \lesssim 10^{15} M_\odot$ . Ils ont trouvé :

$$\Upsilon_{200} \propto M_{200}^{0.26 \pm 0.04} \quad (6.17)$$

soit encore, si on inverse la relation entre  $L$  et  $M$

$$\Upsilon_{200} \propto L^{0.35 \pm 0.02} \quad (6.18)$$

Cette dépendance de  $\Upsilon$  est *faible* par rapport à  $L$ . C'est un comportement auquel on s'attend car la luminosité en bande  $K_S$  est un traceur presque parfait de la masse stellaire. Néanmoins, ici il s'agit de la masse du halo de matière noire, et il est déjà remarquable d'avoir une pente aussi faible entre la masse du halo et la luminosité de ses galaxies.<sup>4</sup> On a essayé, quoique crûment, d'estimer ce rapport

<sup>4</sup>Le  $\Upsilon$  en bande B est plutôt  $\Upsilon \propto L$  pour les grandes luminosités, la dépendance est donc bien moindre ici.

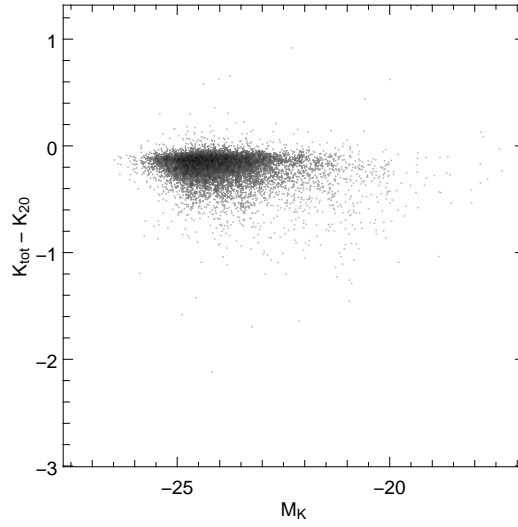


FIG. 6.13: 2MASS Redshift Survey : Comparaison de la magnitude apparente totale, estimée par extrapolation, à la magnitude  $K_{20}$ .

$M/L$  dans le catalogue 2MRS à l'aide du théorème du Viriel. Pour un groupe de galaxies relaxé gravitationnellement, Bahcall & Tremaine (1981); Heisler et al. (1985) ont montré que

$$M_{\text{proj}} = \frac{f_{\text{PM}}}{\pi G(N - \gamma)} \sum_i s_i (V_i - V_G)^2 \quad (6.19)$$

est un assez bon estimateur de la masse virielle, avec  $G$  la constante de gravitation,  $s_i$  la distance projetée sur le ciel entre le barycentre du groupe et la galaxie  $i$ ,  $V_i$  la vitesse apparente de la galaxie  $i$ ,  $V_G$  la vitesse apparente du groupe,  $N$  le nombre de galaxies dans le groupe,  $\gamma = 1.5$  et  $f_{\text{PM}} = 10.2$ . On a représenté sur la figure 6.14 le résultat de l'estimation du paramètre  $\Upsilon_{\text{proj}} = M_{\text{proj}}/L$  pour les deux catalogues MDC et HDC. Si on moyenne par intervalles logarithmiques de luminosité les différents  $M/L$  obtenus on obtient une relation qui ressemble beaucoup à celle de Tully (2005), et dans la partie des grandes luminosité à la relation (6.18). On trouve en effet

$$\Upsilon_{\text{proj}} \propto L^\alpha \quad (6.20)$$

avec  $\alpha \simeq 0.40$  dans le cas de MDC. Néanmoins, au vu de la très grande dispersion des deux graphes, il semble qu'une estimation précise de cette relation soit pour le moins hasardeuse. Nous avons par exemple, qu'en bande B, le minimum du rapport  $M/L$  est atteint pour quelques  $10^{10} L_\odot$ . Hors ici, on trouve plutôt un minimum à  $\sim 10^{12} L_\odot$ . Il s'agit probablement d'un effet de sélection sur les groupes de galaxies détectables à l'aide du 2MRS, ce qui entraîne un changement de pente artificiel. Nous nous contenterons pour cette première approche que nous allons présenter ici, de considérer que le **rapport  $M/L$  est constant** quitte à raffiner cette procédure dans le futur en utilisant l'une de ces relations dont nous venons de parler. On sait par les tests sur catalogues virtuels qu'un tel choix risque de provoquer un biais : les vitesses reconstruites seront sous-estimées. Malgré cela, étant donné l'état des données actuelles il semble que c'est une des meilleures approximations que l'on puisse faire pour le moment.

On peut estimer le rapport  $M/L$  moyen de l'Univers dans la bande  $K$  à l'aide de la relation suivante :

$$M/L = \frac{\Omega_m h^2 \rho_{c,0}}{\int_0^{+\infty} L \Phi(L) dL} = \Omega_m h^2 \rho_{c,0} \left( L_{*,0} \Phi_{*,0} h \int_0^{+\infty} u^{1+\alpha} e^{-u} du \right)^{-1} = 472 \Omega_m h \frac{M_\odot}{L_\odot}, \quad (6.21)$$

avec  $L_{*,0}$  et  $\Phi_{*,0}$  les deux constantes de la fonction de Schechter définie plus haut dans le chapitre mais évaluée pour  $h = 1$ .



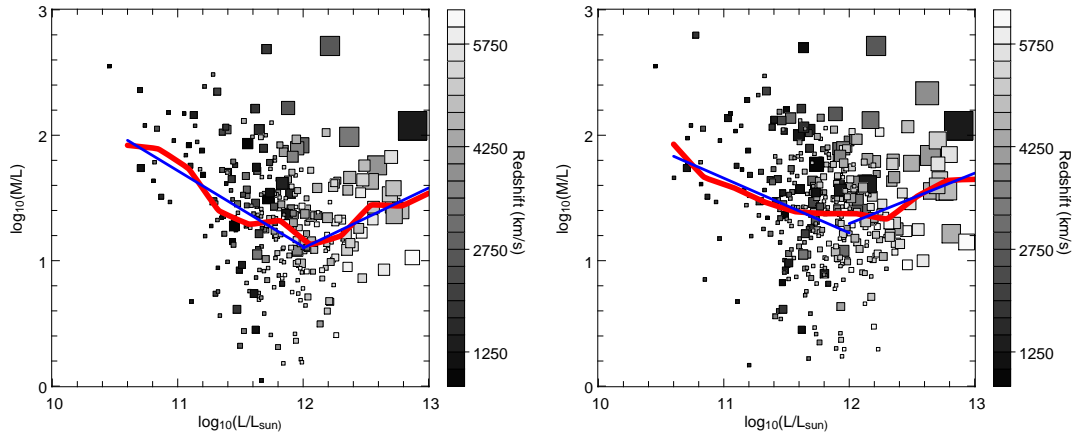


FIG. 6.14: 2MASS Redshift Survey : Estimation du rapport  $M/L$  pour les deux catalogues HDC (à gauche) et MDC (à droite). Les courbes moyennées sont données en rouge sur le graphe. Les meilleurs ajustements par morceaux sont donnés en bleu. On a choisi la luminosité  $10^{12} L_{\odot}$  comme point de changement de la pente de la relation  $M/L$ . Les meilleurs ajustements donnent (avec  $\alpha$  la pente de la droite bleue) :  $\alpha = -0.60$  (pour  $L < 10^{12} L_{\odot}$ ) et  $\alpha = 0.479$  (pour  $L > 10^{12} L_{\odot}$ ) dans le cas de HDC,  $\alpha = -0.43$  (pour  $L < 10^{12} L_{\odot}$ ) et  $\alpha = 0.40$  (pour  $L > 10^{12} L_{\odot}$ ) dans le cas de MDC. Seul les groupes de galaxies ont été représentés ici, ce qui explique que leur nombre semble augmenter quand  $D_0$  augmente.

#### 6.2.4 Préparation du catalogue : phase II

La première phase consistait à obtenir une bonne estimation de la luminosité et de la masse des objets du 2MRS.Rec. Il nous faut à présent nous assurer que l'échantillonnage est homogène : c'est-à-dire il nous faut corriger l'absence de données dans le plan galactique et les effets d'incomplétude. Ces effets ont déjà été discutés dans les sections 5.2.3 (page 76) et 5.4.1 (page 86) respectivement. L'introduction de la masse diffuse, qui fait l'objet de la section 5.2.1 (page 70), dans le catalogue 2MRS fait aussi l'objet de cette section.

##### Remplissage du plan galactique

Nous allons maintenant tenter de nettoyer le catalogue 2MRS des effets d'obscurcissement par le plan galactique (cf. figure 6.4). Sur la figure 6.15, on a représenté la densité de luminosité  $\sigma_L(b)$ , la luminosité présente dans un intervalle de  $\cos b$ , en fonction de la latitude galactique dans le catalogue 2MRS. Si l'échantillonnage était parfaitement isotrope on s'attendrait à ce que cette fonction soit constante. Mais à cause de l'obscurcissement on a une chute brutale de la luminosité dans l'intervalle  $-5 \text{ deg} < b < 5 \text{ deg}$ . Du fait de la présence du bulbe galactique dans la direction  $-30 \text{ deg} < l < 30 \text{ deg}$ , on a un obscurcissement un peu plus fort. La prudence nous inciterait à masquer une région plus importante dans cette direction, par exemple  $-10 \text{ deg} < b < 10 \text{ deg}$ .

Contrairement à l'algorithme de la section 5.4.1, nous n'allons pas ici diviser la masse des galaxies par deux. Mais nous allons dupliquer aléatoirement un objet sur deux de part et d'autre des surfaces  $b = \pm 5 \text{ deg}$  (ou  $b = \pm 10 \text{ deg}$  dans la direction du bulbe galactique).<sup>5</sup> Pour le moment, nous montrons sur la figure 6.16 comment l'algorithme de la section 5.4.1 a été en mesure de remplir le plan galactique. Il semble que l'algorithme joue bien son rôle dans la plupart des directions, où il semble que les structures soient prolongées à travers le plan. Il faut néanmoins nuancer cette satisfaction en regardant dans la direction du super-amas Perseus-Pisces, où la procédure de clonage a semble-t-il rajouté

<sup>5</sup>Cette modification est particulièrement utile dans les cas où nous ne faisons pas des reconstructions avec la résolution maximale. Diviser la luminosité par deux (et donc la masse) introduirait alors un biais dans le maillage MAK dans le volume obscurci par le plan galactique.



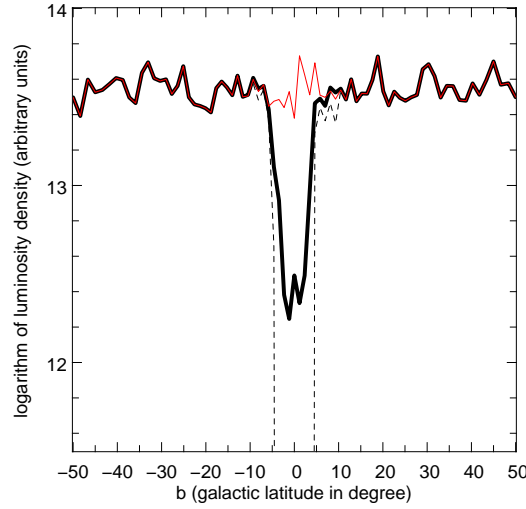


FIG. 6.15: Remplissage du plan galactique – On représente ici la densité de luminosité en fonction de la latitude galactique. La ligne noire épaisse donne la distribution originale de luminosité des galaxies dont on connaît le redshift. La ligne noire en tirets donne cette même distribution une fois que les galaxies avec  $|b| < 5$  deg ont été retirée. La ligne rouge donne la même distribution une fois que les nouvelles galaxies ont été créées en utilisant l’algorithme décrit dans la section 5.4.1.

artificiellement un “petit” amas dans cette direction. Il se peut donc que certain résultat sur les champs de vitesse soit un peu affecté par cette masse excédentaire.

### Correction de l’incomplétude

Une fois que nous avons une distribution couvrant tout le ciel de galaxies, il faut corriger de l’effet de sélection des galaxies en fonction de leur luminosité apparente (effet appelé incomplétude dans les catalogues). Comme nous avons choisi  $M/L = \text{constante}$  l’algorithme de la section 5.2.3 est grandement simplifié et est rendu indépendant du choix de  $\Omega_m$ , qui, n’agissant que comme une constante multiplication sur l’ensemble de la distribution de masses, ne changera pas le champ de vitesse reconstruit. Nous appellerons cette correction une “correction classique”. Il y a néanmoins deux effets supplémentaires qu’il faudrait considérer.

**Estimer la masse manquante** Comme nous l’avons vu à la section 5.2.1, la masse manquante peut avoir un impact significatif sur le champ de vitesse. On peut estimer quelle quantité de masse n’est pas tracée par les galaxies disponibles en cherchant la masse du halo de la plus petite galaxie dans le catalogue. Avec le  $M/L$  pris précédemment, on obtient  $M_{\min} = 1.16 \cdot 10^9 M_{\odot}$ . En utilisant les relations dérivées de la loi de Sheth et Tormen, on obtient une fraction, pour  $\Omega_m = 0.258$ , de l’ordre de 50%. Il nous faudra donc tester l’impact d’ajouter des traceurs de masse fictifs disposés dans le catalogue comme on va le voir dans le paragraphe suivant. Nous n’étudierons pas ici l’impact de l’introduction de traceurs distribués de manière homogène dans les catalogues existant.

**Les effets de masse diffuse** Contrairement à la section 5.2.1, l’effet de masse diffuse dépend aussi de la distance à l’observateur. Pour avoir une idée de l’impact de cette quantité inconnue sur le champ de vitesse reconstruit il nous faut pouvoir l’introduire dans le catalogue 2MRS\_Rec. En fait, pour en tenir compte d’une manière optimale, nous allons l’introduire au niveau du maillage de MAK directement en modifiant l’algorithme de la section 3.3.2 comme suit :

- Nous calculons d’abord une estimation de la quantité de masse manquante dans une coquille sphérique  $S_r$ , placée à une distance  $r$  de l’observateur et avec une épaisseur  $dr$ . Cette quantité

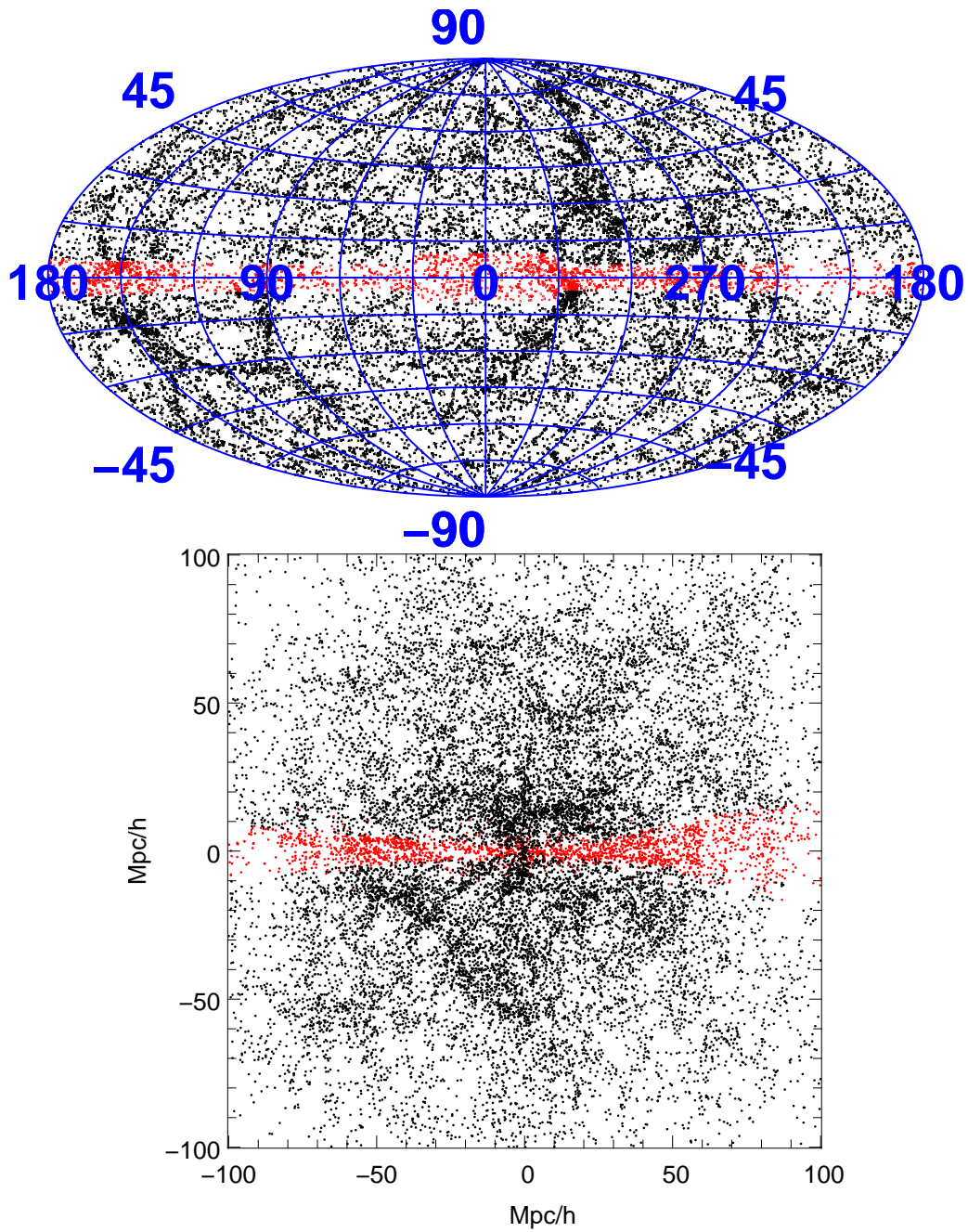


FIG. 6.16: Représentant du catalogue 2MRS après remplissage du plan galactique avec l'algorithme de la section 5.4.1.

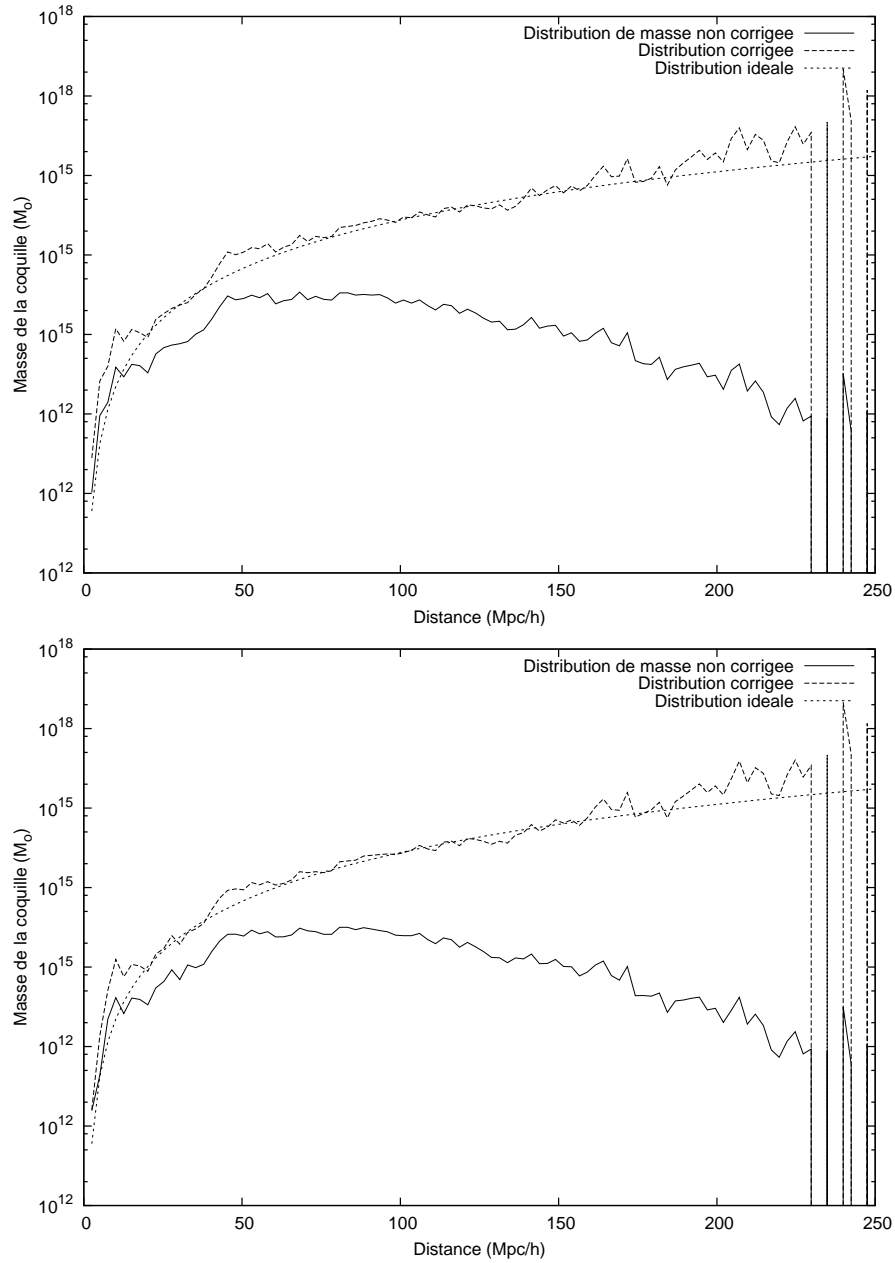


FIG. 6.17: *Correction de l'incomplétude dans le catalogue 2MRS* – Graphe de gauche pour HDC, graphe de droite pour MDC. La courbe idéale, en tirets courts, représente ce qu'on attendrait comme distribution de masse pour un Univers totalement homogène. La ligne solide donne la distribution de masse non corrigée de l'incomplétude. La ligne, en tiring long, donne la distribution de masse après correction de l'incomplétude.

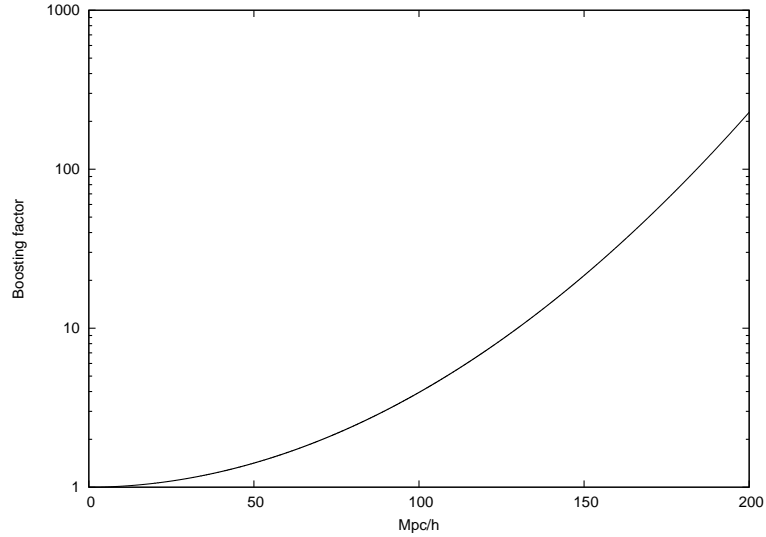


FIG. 6.18: *Correction de l'incomplétude : fonction d'amplification de la luminosité* – On a représenté ici la fonction servant à multiplier la luminosité des objets du catalogue 2MRS à cause de l'extinction naturelle des objets à grand distance. Il s'agit de la fonction  $b(r)$ , dont il est question à la section 6.2.4.

correspond à  $M_{\text{missed},r}$  dans la section 5.2.3, en utilisant la valeur de  $\Upsilon$  spécifié plus haut dans cette section-ci. On construit alors la fraction de masse non-observée dans  $S_r$  :

$$R_{\text{inc}}(r) = \frac{M_{\text{missed},r}}{\Upsilon b(r)(L_{\text{field,obs},r} + L_{\text{groups,obs},r})}. \quad (6.22)$$

La masse  $dm_{\text{inc}}$  qui doit être introduite dans  $S_r$  à cause des effets d'incomplétude/masse diffuse est

$$dm_{\text{inc}}(r) = R_{\text{inc}}(r)\rho_0 4\pi r^2 dr = \rho_0 \mathcal{P}_u(r) dr \quad (6.23)$$

avec  $\rho_0$ , la densité moyenne de l'Univers. Nous sommes maintenant capable de calculer la fraction totale de masse qui doit être introduite dans le catalogue à cause des effets d'incomplétude :

$$f_h = \frac{\int_0^R dm_{\text{inc}}(r)}{\int_0^R \rho_0 4\pi r^2 dr} = \frac{3}{R^3} \int_0^R R_{\text{inc}}(r) r^2 dr \quad (6.24)$$

En utilisant ces équations, on peut calculer le nombre  $N_u$  d'éléments du maillage de MAK qui ne doivent être associé à aucun objet du catalogue  $N_u = N f_h$ .

- Nous pouvons maintenant exécuter l'algorithme de la section 3.3.2 en supposant que le nombre d'éléments du maillage disponibles pour les objets du catalogue est  $N_{\text{obj}} = N(1 - f_h)$ .
- Les éléments du maillage peuvent être distribués aléatoirement dans le volume du catalogue mais en faisant en sorte que la loi de distribution radiale de ces éléments suit  $\mathcal{P}_u(r)$ .<sup>6</sup>

Dans la section 6.3, nous utiliserons cette deuxième correction (que nous nommons lisse) afin de tester la robustesse des résultats obtenus à l'aide de la correction classique. Nous fixerons pour cela une amplification de luminosité maximale ( $b_{\text{max}}$ ), ce qui nous donne une masse manquante  $M_{\text{missed},r} = (b(r) - b_{\text{max}})\Upsilon L_{\text{obs},r}$ . On a donc, dans ce cas simplifié,

$$R_{\text{inc}}(r) = 1 - b_{\text{max}}/b(r). \quad (6.25)$$

<sup>6</sup>Dans le cas de 2MRS-Rec, cette fonction est nulle car  $M_{\text{missed},r} = 0$ , en effet la masse non-observée a déjà été mise dans les galaxies de champ.

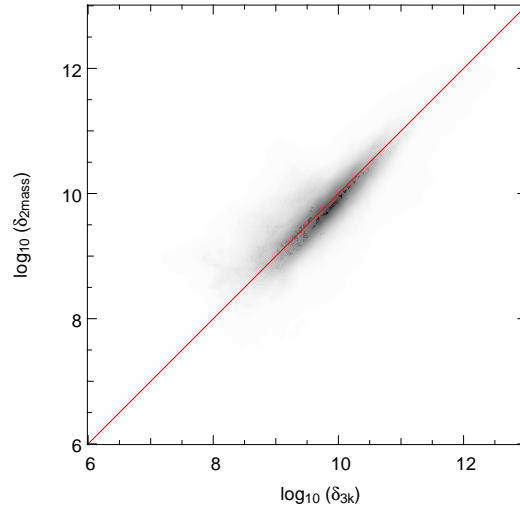


FIG. 6.19: *Comparaison du champ de densité du NBG-3k à celui du 2MRS* – Diagramme de dispersion donnant le résultat de la comparaison du logarithme du champ de densité *non normalisé* du NBG3k au logarithme du champ de densité du catalogue 2MRS obtenu après les corrections des sections 6.2.3, 6.2.3, 6.2.3, 6.2.4 et 6.2.4. Le champ de densité a été obtenu par filtrage adaptatif sur la distribution de masse donné par les catalogues respectifs. On a placé les objets à leur position dans l’espace des redshifts, exprimés dans le référentiel de repos du CMB. Le champ de densité est exprimé en unités arbitraires mais consistant entre les deux catalogues.

Il faut bien remarquer que cette relation est beaucoup plus particulière que celle donnée par l’équation (6.22). Elle suppose intrinsèquement que le rapport  $M/L$  est le même pour *tous* les objets du catalogue. Mais elle a l’avantage d’être très simple.

### 6.2.5 Comparaison du champ de densité du NBG-3k et du 2MASS

On a représenté, figure 6.19, le résultat de la comparaison du logarithme du champ de densité non normalisé du NBG3k au logarithme du champ de densité du catalogue 2MRS. Cette dispersion montre que les deux catalogues sont en très bon accord malgré l’origine très différente de leurs données et de l’affectation du rapport  $M/L$ . On doit s’attendre donc à ce que le catalogue soit une bonne extension du catalogue NBG-3k et permettrait d’étudier le champ de vitesse de manière plus précise en réduisant les effets de bord.

### 6.2.6 Champ de vitesse reconstruit

Nous avons représenté sur les figures 6.20 et 6.21, les vitesses propres des objets du catalogue MDC (obtenu par l’intermédiaire de la reconstruction sur le 2MRS-Rec). Il s’agit d’une vue en projection sur le plan supergalactique ( $SGZ=0$ ), mais incluant les objets situés entre  $-20 h^{-1}\text{Mpc}$  et  $+20 h^{-1}\text{Mpc}$ . On constate que, comme on s’y attendait, les gros amas ont bien tendance à attirer à eux les galaxies qui les entourent. On peut par exemple citer Pisces-Perseus (du côté  $SGX > 0$ ) et Hydra-Centaurus ( $SGX < 0$ ,  $SGY > 0$ ). Un autre amas qui semble attirer à lui des galaxies est Coma. Néanmoins l’alignement de galaxies qu’on voit sur cette projection est probablement fortuit et dû à la projection (en effet en se limitant à une tranche “fine” de  $\pm 10 h^{-1}\text{Mpc}$  le “filament” disparaît). On constate aussi que la densité de galaxies apparente des super-amas décroît qualitativement si on passe en coordonnées réelles (figure 6.21) par rapport au même diagramme formé à partir des galaxies à leur position redshift (figure 6.20, par exemple la région du super-amas Perseus-Pisces). C’est un effet attendu car les positions redshifts sont affectées par les vitesses particulières des objets qui ont tendance à faire se

regrouper des structures spatialement séparées.

Sur la figure 6.22, nous avons représenté le champ de densité en couleur avec en surimpression le champ de vitesse. Ces deux champs ont été obtenus par filtrage adaptatif sur grille fine. On a sous-échantillonné le champ de vitesse d'un facteur quatre dans chaque direction de l'espace afin de pouvoir le représenter clairement.

## 6.3 La vitesse d'ensemble locale et l'origine du dipôle du CMB

### 6.3.1 Mesurer la vitesse d'ensemble : le méthode $\chi^2$

Un problème important de la cosmologie dans l'Univers local est la mesure de la vitesse d'ensemble local (en anglais le “bulk flow”). Cette vitesse correspond, dans la théorie linéaire, à moyenner le champ de vitesse sur le volume considéré. Nous allons voir ici une méthode qui n'utilise pas explicitement cette pondération, mais que, si l'échantillon est assez uniforme donne une bonne approximation de la moyenne volumétrique. Néanmoins, nous verrons qu'il est possible que cet estimateur diffère dans certains cas assez significativement de la moyenne du champ de vecteur *tridimensionnel* :

$$\mathbf{V}_{\text{bulk}} = \frac{1}{V_R} \int_{B_R} d\mathbf{r} \mathbf{v}(\mathbf{r}) \quad (6.26)$$

avec  $V_R$  le volume de la boule  $B_R$  de rayon  $R$  et  $\mathbf{v}(\mathbf{r})$  le champ de vitesse.

Un problème important est que, dans les observations, nous n'avons accès qu'à la composante radiale du champ de vitesse. Nous allons tenter d'estimer le “bulk flow” tridimensionnel avec uniquement la vitesse radiale. Cela nécessite implicitement une extrapolation et il est possible, qu'à cause de configurations spécifiques des objets, il existe un biais local énorme vis à vis du vrai “bulk flow”  $\mathbf{V}_{\text{bulk}}$ .

#### Théorie (sur objet ponctuels)

Nous allons utiliser l'estimateur suivant pour le bulk flow :

$$\chi^2(H, \mathbf{U}) = \sum_{i=1}^N \frac{(V_i - H_0 d_i - U^k \hat{x}_i^k)^2}{\sigma_i^2}, \quad (6.27)$$

avec  $V_i = cz_i$  la vitesse apparente de l'objet  $i$ ,  $d_i$  sa distance mesurée,  $U^k$  le “bulk flow”,  $\hat{x}_i$  le vecteur unitaire pointant dans la direction de l'objet  $i$ ,  $\sigma_i$  l'erreur portant sur l'estimation du bulk flow par l'intermédiaire de l'objet  $i$ ,  $N$  le nombre d'objet pour l'estimation de ce bulk flow. Si on minimise la quantité (6.27) par rapport à  $\mathbf{U}$ , on obtient l'équation suivante :

$$-\frac{1}{2} \frac{\partial \chi^2}{\partial U^j} = \sum_{i=1}^N \hat{x}_i^j \frac{(V_i - H_0 d_i - U^k \hat{x}_i^k)}{\sigma_i^2} \quad (6.28)$$

$$= A^j - M^{j,k} U^k \quad (6.29)$$

$$= 0 \quad \text{pour trouver le minimum} \quad (6.30)$$

avec

$$A^j = \sum_{i=1}^N \frac{\hat{x}_i^j (V_i - H_0 d_i)}{\sigma_i^2} \quad (6.31)$$

$$M^{j,k} = \sum_{i=1}^N \frac{\hat{x}_i^j \hat{x}_i^k}{\sigma_i^2} \quad (6.32)$$

On a donc :

$$U^k = (M^{-1})^{k,j} A^j \quad (6.33)$$

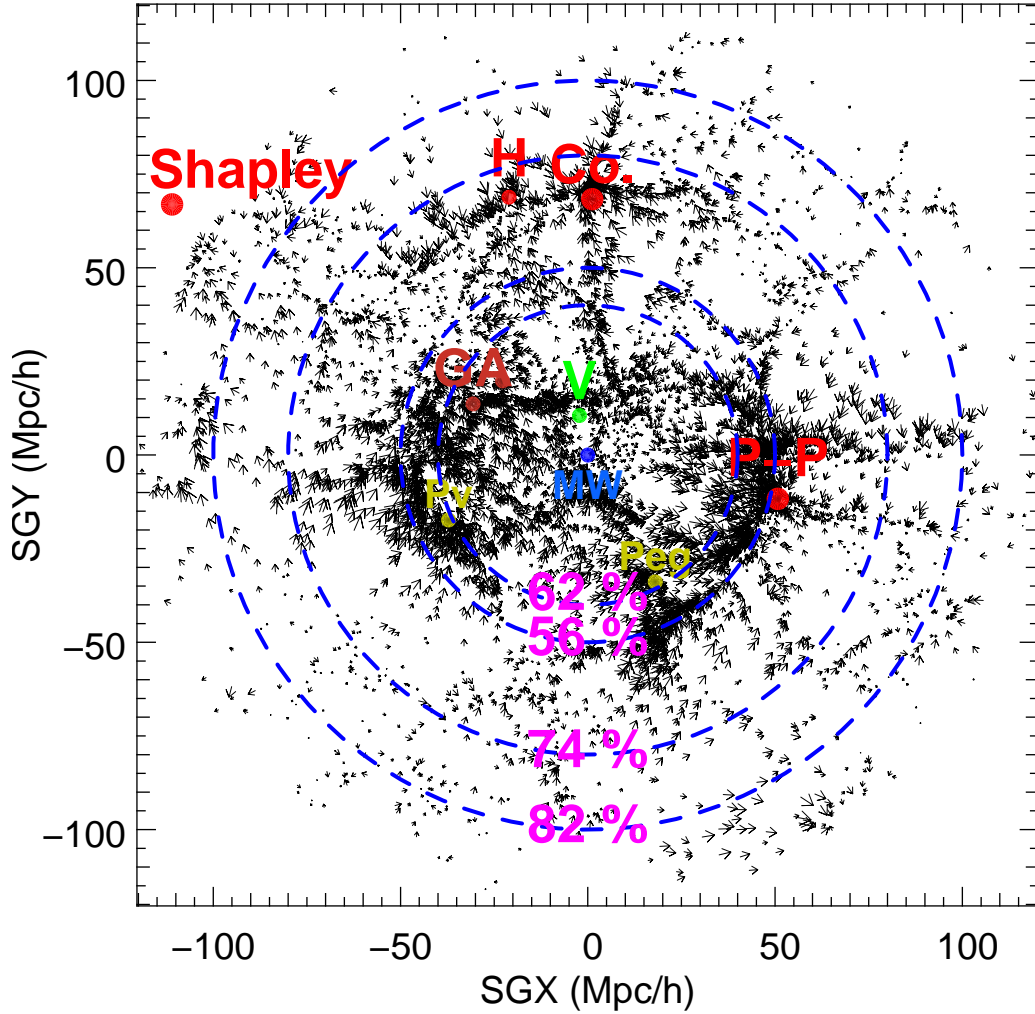


FIG. 6.20: *Vue projetée des vitesses propres obtenues par la reconstruction de  $120 h^{-1} \text{Mpc}$  de profondeur du catalogue 2MRS/MDC. La projection est épaisse de  $40 h^{-1} \text{Mpc}$  et centré sur le plan supergalactique  $\text{SGZ}=0 \text{ km s}^{-1}$ . On a rajouté l'emplacement de structures connues : la Voie Lactée ("MW" en bleu), Virgo ("V" en vert clair), Hydra/Centaurus ("GA" en rouge), les amas Pegasus et Pavo ("Pv" et "Peg" en jaune gris), Coma ("Co." en haut à droite en rouge), Pisces-Perseus (en bas à droite en rouge), Hercules (en haut à gauche en rouge). Les galaxies ont été placées à leur position donnée par le redshift (voir figure 6.21 pour la position réelle reconstruite). On a représenté à l'aide de cercle concentrique bleu les différentes étapes observées de la convergence de l'amplitude du mouvement du Groupe Local. Le pourcentage indiqué est pris par rapport à la meilleure estimation provenant de l'étude du fond diffus cosmologique.*

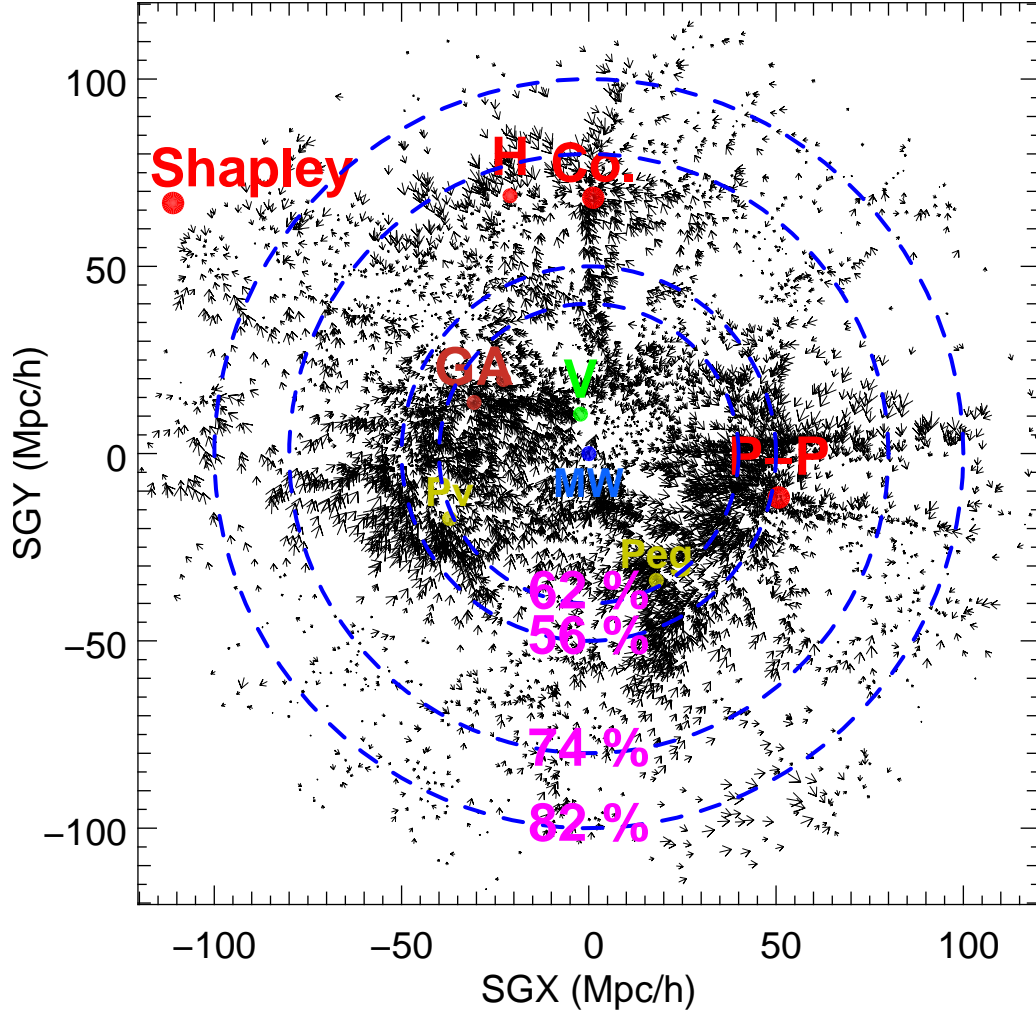


FIG. 6.21: *Vue projeté des vitesses propres obtenues par la reconstruction de  $120 h^{-1} \text{Mpc}$  de profondeur du catalogue 2MRS/MDC – Même chose que dans la figure 6.20 mais les objets ont été placé à leur position réelle reconstruite.*



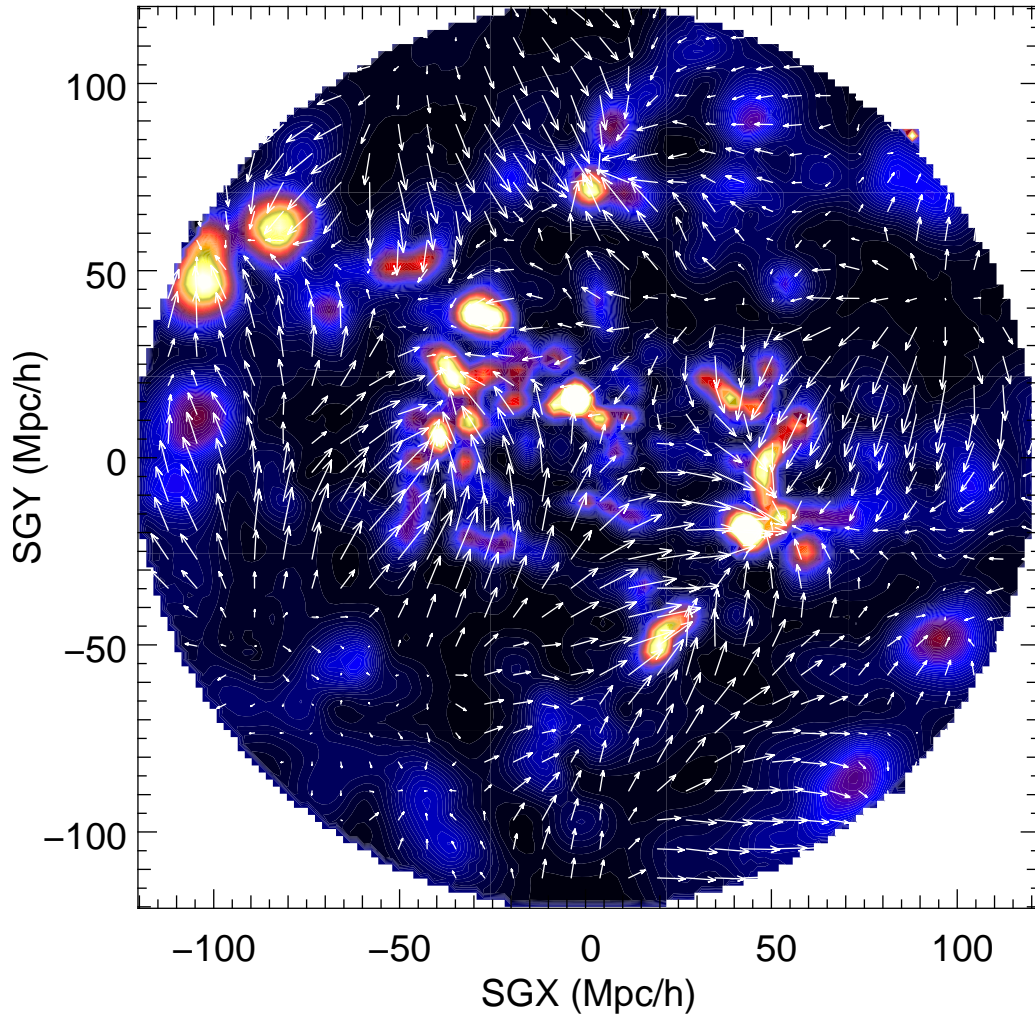


FIG. 6.22: Illustration du champ de vitesse obtenu par lissage adaptatif sur les vitesses reconstruites par la méthode de reconstruction MAK. Le champ de densité correspondant a été représenté avec un codage de couleur en dessous le champ de vitesse. On a choisi ici une coupe infiniment fine du champ selon le plan supergalactique  $SGZ = 0$ . Le code de couleur est proportionnel au logarithme de la densité local : le bleu représente des régions sous-denses et le rouge des régions sur-denses.

où  $M$  est la matrice de covariance de l'échantillon considéré. On peut alors trouver une estimation de l'erreur statistique  $\sigma_{U^k}$  sur chacune des composantes  $U^k$  (Press et al., 1992) :

$$\sigma_{U^k}^2 = \sum_{i=1}^N \sum_{l=1}^3 \sigma_i^2 \left( \frac{(M^{-1})_{k,l} \hat{x}_i^l}{\sigma_i^2} \right)^2. \quad (6.34)$$

Nous voudrions à présent savoir si cet estimateur est un bon estimateur du bulk flow dans la simulation.

### Théorie (sur champ de vitesse)

On peut créer un estimateur équivalent que précédemment mais en supposant que l'on a accès au champ de vitesse radial complet (et donc qu'il n'y a pas de problème d'échantillonnage, contrairement à ce qui est discuté par exemple dans Kaiser (1988); Regos & Szalay (1989)). Pour cela, nous minimiserons la quantité :

$$\chi^2(H, \mathbf{U}) = \int_{B_R} d^3\mathbf{r} \left( v_r(\mathbf{r}) - \sum_i U_i \hat{\mathbf{r}}_i \right)^2, \quad (6.35)$$

qui est l'équivalent "continu" de la quantité (6.27), avec  $B_R$  la boule de rayon  $R$  centré sur l'observateur et  $v_r(\mathbf{r}) = \hat{\mathbf{r}}_k v_k(\mathbf{r})$  la composante du champ de vitesse selon la ligne de visée. Si on calcul ce qu'est le vecteur  $\mathbf{U}$  satisfaisant la condition de minimalité  $\partial\chi^2/\partial U_i = 0$ , on trouve :

$$-\frac{1}{2} \frac{\partial\chi^2}{\partial U_i} = \int_{B_R} d^3\mathbf{r} \hat{\mathbf{r}}_i \left( v_r(\mathbf{r}) - \sum_j U_j \hat{\mathbf{r}}_j \right) \quad (6.36)$$

$$= \left( \int_{B_R} d^3\mathbf{r} v_r(\mathbf{r}) \hat{\mathbf{r}}_i \right) - \sum_j U_j \int_{B_R} d^3\mathbf{r} \hat{\mathbf{r}}_j \hat{\mathbf{r}}_i \quad (6.37)$$

$$= \left( \int_{B_R} d^3\mathbf{r} v_r(\mathbf{r}) \hat{\mathbf{r}}_i \right) - U_i \frac{V}{3} \quad (6.38)$$

on a donc

$$U_i = \frac{3}{V} \int_{B_R} d^3\mathbf{r} v_r(\mathbf{r}) \hat{\mathbf{r}}_i = \frac{3}{V} \sum_k \int_{B_R} d^3\mathbf{r} \hat{\mathbf{r}}_k v_k(\mathbf{r}) \hat{\mathbf{r}}_i. \quad (6.39)$$

On peut remarquer qu'en fait la matrice de covariance est totalement triviale ici. Cela est dû au fait que l'échantillonnage étant bien régulier il n'y a plus de corrélation entre les différentes directions.

### Test sur des simulations

Pour cela on a utilisé une simulation effectuée dans une boîte de  $500 h^{-1}\text{Mpc}$ , avec  $256^3$  particules. Cette simulation a été effectuée comme précédemment au chapitre 3 (page 29) avec la version publique du code HYDRA (Couchman et al., 1995). On a choisi des paramètres cosmologiques habituels pour une cosmologie  $\Lambda\text{CDM}$ . La densité moyenne de matière est  $\Omega_m = 0.30$  et la constante cosmologique  $\Omega_\Lambda = 0.70$ . La constante de Hubble est  $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . La normalisation des fluctuations de densité dans une sphère de rayon  $8 h^{-1}\text{Mpc}$  est  $\sigma_8 = 1.0$ . Une coupe du champ de densité, obtenu par filtrage adaptatif, de cette simulation est représentée figure 6.23. A partir de cette échantillon, on a fabriqué un nouveau catalogue de halos en utilisant un algorithme friend-of-friend. On appelle ce catalogue *FullMock500*. La fraction de masse dans les halos est environ 37%, on a donc encore une fraction de 63% de la masse laissée dans un champ d'arrière-plan non effondré dans des structures compactes. De *FullMock500*, on a tiré aléatoirement dix observateurs. A partir de ces observateurs nous avons calculé une succession de bulk flows pour des volumes sphériques de rayon  $R$  croissant. Ces bulk flows ont été calculé de quatre manières différentes :

- I. en utilisant les groupes et galaxies de champ de FullMock et en appliquant la méthode du paragraphe précédent. Pour un rayon  $R$  donné, on sélectionne les groupes se trouvant à une vraie distance de l'observateur inférieure à  $R$  et on utilise la méthode ci-dessus sur ce sous-ensemble.
- II. en calculant la moyenne simple des vitesses tridimensionnelle des objets du catalogue présents dans le volume de rayon  $R$ .
- III. en effectuant au préalable un filtrage adaptatif sur une grille fine ( $256^3$ ) du champ de vitesse tridimensionnel. On calcule ensuite le bulk flow en utilisant le champ de vitesse calculée sur une grille régulière pour évaluer l'intégrale (6.26) sur la boule de rayon  $R$ . à l'intérieur d'une sphère de rayon  $R$  et centré sur l'observateur.
- IV. en calculant, avec ce même champ de vitesse issu du filtrage adaptatif, une estimation du bulk-flow en utilisant l'équation (6.39).

Les résultats donnés par ces deux approches sont donnés sur la figure 6.24 pour les quatre méthodes et les dix observateurs. Seule l'amplitude du bulk flow a été tracée. On voit néanmoins déjà que, bien que dans l'ensemble les quatre courbes se ressemblent, il y a de très fortes différences dans le détail. Après examen, on constate :

- les méthodes II et III donnent quasiment le même résultat,
- les méthodes I et IV donnent aussi quasiment le même résultat,
- les quatre méthodes donnent en moyenne les mêmes tendances mais ont de fortes disparités particulières.

Il semble que les différences entre (I & IV) et (II & III) proviennent purement d'un effet de projection sur la ligne de visée. Si par malheur, une grosse partie du bulk flow provient d'une région de l'espace où la vitesse est orthogonale à la ligne de visée, on n'observera pas cet effet dans le bulk flow estimé à partir des vitesses radiales. Si on fait des comparaisons entre les bulk flow il faudra donc impérativement utiliser la même méthode de mesure pour être sensible au même biais. D'un autre côté, cela montre qu'on a effectivement besoin de méthodes de reconstruction du champ de vitesse pour être véritablement en mesure de l'exploiter correctement. Les vitesses radiales observées ne suffisent même pas pour être sûr qu'il ne nous manque pas une composante importante du bulk flow du champ de vitesse. Cet effet avait déjà été indiqué par Regos & Szalay (1989) mais ils n'avaient pas insisté sur l'amplitude exacte de la variance additionnelle qui est introduite par la projection. Regos & Szalay (1989) (en conjonction avec l'annexe C.3) nous permettent d'écrire que les variances de  $U_i$  et  $V_{\text{bulk},i}$  sont toutes deux données par

$$\frac{(\beta H)^2}{6\pi^2} \int dk P(k) W_R(k)^2 \quad (6.40)$$

avec, grossièrement si on essaye de mesurer le bulk-flow avec le champ de vitesse,

$$W_R(k) = \begin{cases} \frac{3}{(kR)^3} (\sin(kR) - kR \cos(kR)) & \text{pour } V_{\text{bulk},i} \\ \frac{3}{(kR)^3} \int_0^{Rk} du u^2 (j_0(u) - 2j_2(u)) & \text{pour } U_i \end{cases} \quad (6.41)$$

Dans l'annexe C.3, on trouve que on peut avoir une variance supplémentaire de  $\sim 80 \text{ km s}^{-1}$  sur chaque composante du bulk flow, à une distance de  $40 h^{-1} \text{ Mpc}$ , par l'utilisation de l'estimateur basé sur les vitesses radiales. Cela se traduit par une variance  $\sim 140 \text{ km s}^{-1}$  sur l'amplitude. C'est une variance énorme qui explique les différences entre les courbes de la figure 6.24.

### 6.3.2 Analyse de la vitesse d'ensemble dans une simulation $\Lambda \text{CDM}$

Nous utiliserons ici la simulation (et le catalogue virtuel de base FullMock500) introduite à la section 6.3.1. Parmi les groupes identifiés dans FullMock500, on souhaite en isoler un qui soit dans un environnement comme le Groupe Local. On en a trouvé dont les caractéristiques environnementales sont les suivantes :

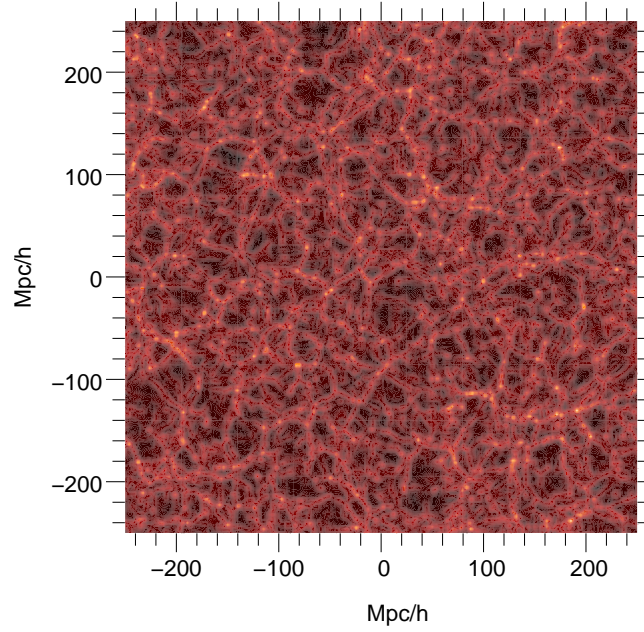


FIG. 6.23: *Test de la procédure sur une simulation* – Ceci est représentation d’une coupe du champ de densité de la simulation de  $500 h^{-1}\text{Mpc}$  de côté telle qu’elle est décrit dans le text principal (section 6.3.2). L’échelle de couleur (noir au rouge brillant) est proportionnelle au logarithme de la densité.

- il s’agit d’un “petit” groupe de masse  $4.9 \cdot 10^{12} h^{-1} M_{\odot}$  ;
- sa vitesse est de  $v = 603 \text{ km s}^{-1}$ , dans la direction  $l = 256$ ,  $b = 52$  ;
- un Univers local assez vide, avec une densité relative dans un volume de rayon  $20 h^{-1}\text{Mpc}$ ,  $30 h^{-1}\text{Mpc}$  et  $100 h^{-1}\text{Mpc}$  respectivement de  $\delta_{20} = -0.43$  ( $\Omega_{\text{m,eff}} = 0.171$ ),  $\delta_{30} = -0.001$  ( $\Omega_{\text{m,eff}} = 0.30$ ),  $\delta_{100} = -0.028$  ( $\Omega_{\text{m,eff}} = 0.29$ ).

Ayant choisi ce groupe de particules comme groupe local virtuel, on place l’observateur à cet endroit et on construit deux catalogue virtuels de volume fini de rayon  $100 h^{-1}\text{Mpc}$ , sans incomplétude et dont les coordonnées redshifts seules sont connues, sauf à l’intérieur d’une sphère de  $30 h^{-1}\text{Mpc}$  de rayon où l’on connaît aussi les vitesses propres. Le premier catalogue, Mock2MRS-CMB, utilise les coordonnées redshifts exprimées dans le référentiel de repos de la simulation entière (d’où le suffixe CMB). Le second catalogue virtuel, Mock2MRS-LG, est identique au précédent mis à part que les redshifts sont exprimés dans le référentiel de repos du groupe local virtuel. Une fois ces deux catalogues construits, on effectue sur chacun une reconstruction PaddedDom (§ 5.4.2, page 87) tenant compte du décalage vers le rouge. Contrairement aux reconstructions du chapitre 5, on supprime ici les objets ayant un décalage vers le rouge négatif. Nous considérerons que les masses sont connues exactement et que l’incomplétude n’ajoute aucun effet (ce qui a été vu à la section 5.2.3, page 76).

Nous avons essayé de calculer la vitesse d’ensemble dans deux référentiels différents :

- d’abord nous avons choisi le référentiel du “groupe local”, tous les décalage vers le rouge du catalogue virtuel sont exprimés à l’aide des vitesses des objets dans le référentiel du “groupe local” (LG par la suite),
- ensuite nous nous sommes placés dans le référentiel de repos de la simulation, que nous appelons “CMB”.

La reconstruction est exécutée en supposant  $\Omega_{\text{m}} = 0.30$ . On interpole, à l’aide du filtre adaptatif de l’annexe E (page 175), le champ de vitesse au centre du volume de reconstruction, où l’observateur se trouve en principe. Les vitesses trouvées pour les deux reconstructions sont données table 6.1. Cette vitesse correspond au mouvement de grande échelle d’une région située à la place du “LG”. Vu que les objets dans cette région ont décalage vers le rouge négatif et que la reconstruction MAK ne fonctionne

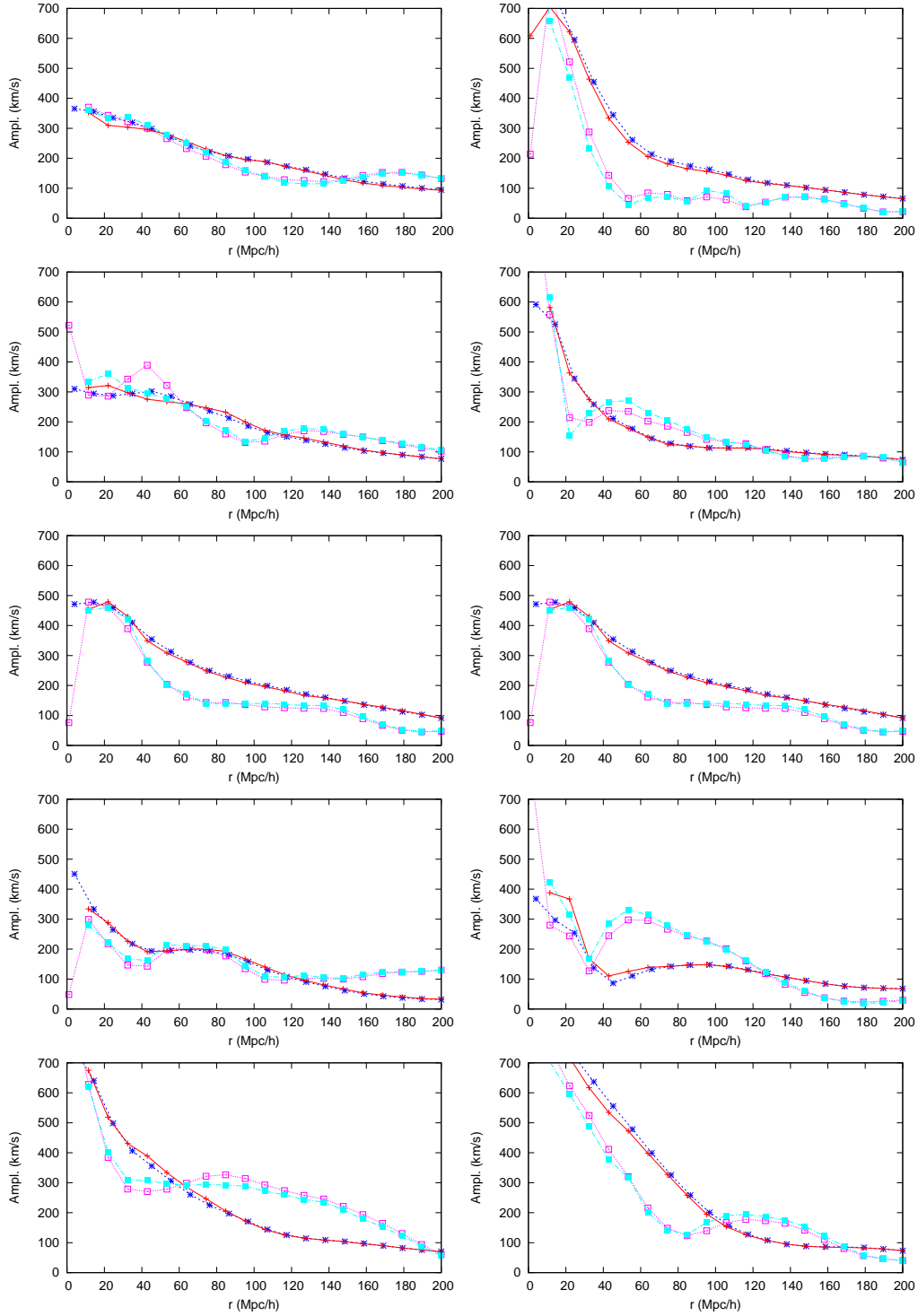


FIG. 6.24: *Estimation de bulk-flow dans une simulation* – On représente ici l'évolution de l'amplitude du bulk flow pour un volume croissant de volume considéré. On a considéré 10 observateurs pris aléatoirement dans le volume de la simulation ( $500^3 h^{-3} \text{Mpc}^3$ ). L'amplitude a été mesurée de deux manières différentes. Les courbes en bleu sombre donnent l'amplitude mesurée à partir de la moyenne volumique du *champ tridimensionnel* de vitesse de la simulation. Les courbes en rouge donnent l'amplitude mesurée à partir de la moyenne des vecteurs vitesses des traceurs de la simulation. Les courbes en cyan correspondent à l'amplitude du vecteur vitesse bulk-flow mesurée à l'aide de la composante sur la ligne des visée des vitesses des traceurs. Les courbes en magenta ont été obtenues en mesurant l'amplitude du vecteur vitesse bulk-flow à l'aide de la composante sur la ligne de visée du *champ* de vitesse.

OBSERVATION VIRTUELLE	
Référentiel du “LG”	Référentiel du “CMB”
$V_{\text{LG/LG}} = 0$ (par construction)	$V_{\text{LG/CMB}} = 603 \text{ km s}^{-1}$ , $l = 256, b = 52$
RECONSTRUCTION SUR UN VOLUME DE RAYON $100 h^{-1}\text{Mpc}$	
Dans le référentiel du “LG”	Dans le référentiel du “CMB”
$V_{\text{LG/LG}} = 942 \text{ km s}^{-1}$ , $l = 249, b = 52$	$V_{\text{LG/CMB}} = 838 \text{ km s}^{-1}$ , $l = 294, b = 57$

TAB. 6.1: *Mesure du l'effet dipole du CMB dans un catalogue virtuel* – On donne ici les résultats de la reconstruction appliquée à la simulation pour déterminer la vitesse de l'observateur dans deux référentiels différents.

qu'à grande échelle on s'attend à ce que la vitesse interpolée corresponde aux mouvement d'une masse de la taille du “LG” ou un peu plus gros (un petit groupe de galaxie par exemple), ce qui est la vitesse que l'on cherche effectivement.<sup>7</sup> On constate d'abord que l'amplitude est beaucoup trop élevée dans le référentiel LG. Malheureusement c'est un effet auquel on peut s'attendre car la reconstruction MAK n'est pas forcée d'imposer une vitesse nulle au niveau de l'observateur. Néanmoins cette vitesse sera utile par la suite pour la reconstruction de la vitesse d'ensemble du volume de  $30 h^{-1}\text{Mpc}$  de rayon. Si on se tourne vers la vitesse interpolée au centre pour la reconstruction dans le référentiel du “CMB”, on trouve un vecteur en bon accord avec le mouvement connu du “LG”. En effet, bien que la vitesse soit un peu surestimée (on reconstruit une vitesse de  $838 \text{ km s}^{-1}$  à la place de  $603 \text{ km s}^{-1}$ ), l'angle entre les deux vecteurs est de l'ordre de 22 deg.

Après avoir regardé cet effet à petite échelle, on peut regarder si on est capable de reconstruire le mouvement d'ensemble d'un volume d'Univers simulé de  $30 h^{-1}\text{Mpc}$  de rayon. Ce mouvement d'ensemble est estimé avec la même méthode et sur le sous-ensemble d'objet qui a une mesure de vitesse radiale et qui a été reconstruit. Dans le cas de la reconstruction dans le référentiel du “LG”, on a de plus retranché la vitesse mesurée dans le référentiel du “LG”. Ces résultats sont présentées dans la table 6.2. On constate cette fois dans les deux cas un assez bon accord entre les observations virtuelles par les champ de vitesses “observées” et le résultat donné par la reconstruction. On a un effet un angle  $\alpha = 25 \text{ deg}$  dans le cas d'une reconstruction “CMB” et  $\alpha = 34 \text{ deg}$  dans le reconstruction “LG” entre la vitesse d'ensemble reconstruite et observée. De plus, l'amplitude de ce vecteur vitesse est  $50 \text{ km s}^{-1}$  plus grande dans une reconstruction “CMB” par rapport à l'observation, et  $100 \text{ km s}^{-1}$  plus grande dans le cas d'une reconstruction “LG”. Ces résultat nous rendent confiants de notre capacité à effectuer des mesures crédibles sur les vrais catalogues de galaxies.

### 6.3.3 La vitesse d'ensemble observée par le catalogue NBG-3k

On applique maintenant cette méthode au catalogue NBG-3k. Nous donnons les résultats de reconstructions du mouvement du Groupe Local, à l'aide du champ de vitesse interpolé à sa position comme on a fait précédemment avec le catalogue virtuel, dans la table 6.3. On voit que sa vitesse par rapport au référentiel de la reconstruction (supposé proche du CMB) est moindre que la vitesse observée, néanmoins sa direction est quasiment parfaite avec un écart de 25 degrés entre les deux vecteurs. D'après les résultats obtenus par d'autres groupes (par exemple Erdogdu et al., 2006a; Erdoğan et al., 2006b; Kocevski & Ebeling, 2006), la convergence du champ de gravitation donnant lieu à notre vitesse n'est pas achevée avant d'avoir considéré un volume de  $60 h^{-1}\text{Mpc}$  voire  $200 h^{-1}\text{Mpc}$ . Ce résultat est en parfait accord avec les leurs. Nous donnons aussi les résultats du bulk flow reconstruit et observé de l'ensemble des galaxies à l'intérieur de la sphère de  $3000 \text{ km s}^{-1} \text{ Mpc}^{-1}$  pour une re-

<sup>7</sup>Le filtre adaptatif a utilisé un rayon de lissage  $4.65 h^{-1}\text{Mpc}$ .

OBSERVATIONS VIRTUELLES	
Référentiel du “LG” $V_{\text{LG/LSC}} = 423 \pm 13 \text{ km s}^{-1}$ , $l = 266 \pm 3, b = 41 \pm 2$	Référentiel du “CMB” $V_{\text{LSC/CMB}} = 249 \pm 16 \text{ km s}^{-1}$ , $l = 227 \pm 10, b = 69 \pm 3$
RECONSTRUCTION SUR UN VOLUME DE RAYON $100 h^{-1}\text{Mpc}$	
Dans le référentiel du “LG” $V_{\text{LG/LSC}} = 527 \text{ km s}^{-1}$ , $l = 223, b = 34$ ( $\alpha = 34 \text{ deg}$ par rapport aux observations)	Dans le référentiel du “CMB” $V_{\text{LSC/CMB}} = 296 \pm 16 \text{ km s}^{-1}$ , $l = 307 \pm 3, b = 72 \pm 1$ ( $\alpha = 25 \text{ deg}$ par rapport aux observations)

TAB. 6.2: *Mesure et reconstruction de bulk flow  $3000 \text{ km s}^{-1}$  dans un catalogue virtuel* – présente ici les résultat de l’observation virtuelle de la vitesse d’ensemble d’un volume de  $3000 \text{ km s}^{-1}$  dans deux référentiels différents : celui du “LG” (tel qu’il a été identifié dans la simulation) et celui du “CMB” (le référentiel de la simulation complète). Dans le cas du référentiel du “LG” on a pris soin de soustraire la vitesse reconstruite du “LG” dans son référentiel (afin qu’il soit vraiment au repos) à la vitesse d’ensemble mesuré dans ce même référentiel.

OBSERVATIONS	
Vitesse du Groupe Local par rapport au référentiel du LG : $V_{\text{LG/LG}} = 0$ (par construction)	Vitesse du Groupe Local par rapport au référentiel du CMB : $V_{\text{LG/CMB}} = 627 \pm 22 \text{ km s}^{-1}; l = 276, b = 30$
RECONSTRUCTION À L’AIDE DES DONNÉES REDSHIFT DU NBG-3K	
Vitesse reconstruite interpolée de la matière à la position du Groupe Local par rapport au référentiel LG : $V_{\text{LG/LG}} = 459 \text{ km s}^{-1}; l = 285, b = 38$	Vitesse reconstruite interpolée de la matière à la position du Groupe Local par rapport au référentiel CMB : $V_{\text{LG/CMB}} = 373 \text{ km s}^{-1}; l = 247, b = 37$

TAB. 6.3: *Reconstruction du mouvement du Groupe Local*

construction ayant été faite sur les galaxies comprises dans ce même volume (excepté la Voie Lactée et M31). Ces résultats sont donnés dans la table 6.4. Nous avons choisi un  $\sigma_i = 300 \text{ km s}^{-1}$  pour toutes les vitesses suivant Sarkar et al. (2007). Ce choix n’affecte pas la meilleure estimation mais change l’amplitude de la barre d’erreur.

### 6.3.4 La vitesse d’ensemble reconstruite à partir du catalogue 2MRS

#### Le mouvement du groupe local

La table 6.5 résume les résultats obtenus sur la mesure de la vitesse de la matière à la position du Groupe Local à l’aide d’une reconstruction sur les données du catalogue 2MRS. On constate qu’en effet la vitesse augmente progressivement pour des volumes de reconstructions  $R_{\text{rec}}$  allant de 10 à  $150 h^{-1}\text{Mpc}$ . Nous n’allons pas jusqu’à  $200 h^{-1}\text{Mpc}$  car la courbe de la figure 6.18 montre que l’on serait obligé de multiplier les luminosités par un facteur 200 à cette distance. Cela rend la reconstruction extrêmement sensible au bruit de grenaille des galaxies. En effet, chaque galaxie un peu lumineuse peut apparaître comme un amas avec cette correction. Cela peut faire osciller et/ou

OBSERVATIONS NBG-3K (DISTANCES)	
Référentiel du “LG” $V_{\text{LG/LSC}} = 354 \pm 33 \text{ km s}^{-1}$ ; $l = 241 \pm 8, b = 43 \pm 5$	Référentiel du “CMB” $V_{\text{LSC/CMB}} = 359 \pm 25 \text{ km s}^{-1}$ ; $l = 298 \pm 4, b = 9 \pm 3$
RECONSTRUCTION SUR UN VOLUME DE RAYON $30 h^{-1}\text{Mpc}$ (N BG-3K)	
Dans le référentiel du “LG” $V_{\text{LG/LSC}} = 314 \text{ km s}^{-1}$ ; $l = 214, b = 14$	Dans le référentiel du “CMB” $V_{\text{LSC/CMB}} = 104 \pm 24 \text{ km s}^{-1}$ ; $l = 5 \pm 19, b = 31 \pm 15$

 TAB. 6.4: *Mesure et reconstruction de bulk flow  $3000 \text{ km s}^{-1}$  à l'aide du catalogue NBG-3k*

accélérer le Groupe Local plus que de raison. Si le bruit de grenaille était négligeable, on ne devrait pas avoir de différence entre le résultat obtenu entre un catalogue corrigé avec une méthode standard (en amplifiant isotropiquement la luminosité) et un autre obtenu par la procédure alternative de la page 120 (nous rappelons que nous l'avons appelé correction lisse). En effet celle-ci consiste à placer la masse manquante non pas dans les traceurs existant mais de manière plus homogène. En fait, on s'attend à sur-estimer la variance du champ de densité dans le premier cas et à la sous-estimer dans le deuxième cas.

Sur cette plage de rayons, on constate que l'amplitude de la vitesse du Groupe Local semble converger vers sa valeur que l'on connaît par les mesures du dipole CMB. Néanmoins sa direction est encore incertaine. En effet, pour des petits volumes ( $40 h^{-1}\text{Mpc}$ , croix et carré bleu) la direction de la vitesse est quasiment parfaitement alignée avec celle du CMB (carré noir). Cela correspond à une attraction par l'amas Hydra-Centaurus. On voit que son effet a été d'attirer le Groupe Local dans la direction donnée par la croix et le carré bleu. Néanmoins, quand on ajoute l'amas Norma et Perseus-Pisces (situé sous le plan galactique, cf. figure 2.3, page 7 ou carte à  $50 h^{-1}\text{Mpc}$  de la figure 6.26), on fait tourner la vitesse du Groupe Local vers le triangle bleu. Cela signifie qu'il y a une lutte à grande échelle entre les amas pour produire notre vitesse : très localement il ne semble pas y avoir de problèmes mais à l'échelle supérieure notre vitesse oscille. Si on ajoute Coma et le “Great Wall” (disque majenta) on constate qu'on ramène la vitesse au-dessus du plan galactique mais pas encore assez proche de celle donnée par le CMB. Si on va à des profondeurs plus grandes ( $120 h^{-1}\text{Mpc}$ , croix magenta et  $150 h^{-1}\text{Mpc}$  carré rouge), la vitesse continue d'augmenter avec la méthode de correction classique de la luminosité. Néanmoins la direction donnée par le carré rouge pointe cette fois vers l'hémisphère galactique sud. Cet effet est probablement du au bruit de grenaille comme expliqué ci-dessus, si on utilise la correction lisse on obtient alors les symboles verts. Il n'y a pas beaucoup de différences à  $120 h^{-1}\text{Mpc}$ , mais il y en a beaucoup à  $150 h^{-1}\text{Mpc}$ . La direction donnée par la correction lisse est cette fois bien dans l'hémisphère nord, quoique toujours à  $50$  degrés de la direction donnée par le CMB.

Avec la méthode de construction de 2MRS-Rec, il semble qu'il y ait un écart systématique assez important au-delà de  $80 h^{-1}\text{Mpc}$ . La vitesse reconstruite pointe alors à  $50$  degrés de la direction de la vitesse donnée par le CMB. Si on regarde la projection aitoft, on s'aperçoit que s'il y avait une masse significative dans la direction du bulbe galactique et à  $\sim 80 h^{-1}\text{Mpc}$  il serait peut-être possible de nous amener plus près du dipole. Pour plus de clarté, nous avons représenté, en projection aitoft, des coupes en distance cartes du champ de densité obtenu par lissage adaptatif sur la distribution de masse tracée par les objets du catalogue 2MRS-Rec. Il s'agit de coupe infiniment fine du champ, mais dont le rayon de lissage typique augmente avec la distance. On repère ainsi des structures que l'on connaît bien : à  $17 h^{-1}\text{Mpc}$  (Virgo),  $30 h^{-1}\text{Mpc}$  (Hydra),  $40 h^{-1}\text{Mpc}$  (Centaurus),  $50 h^{-1}\text{Mpc}$  (Perseus-Pisces à gauche, Norma à droite),  $70 h^{-1}\text{Mpc}$  (Coma),  $100 h^{-1}\text{Mpc}$  (Hercules),  $120 h^{-1}\text{Mpc}$  (Shapley à droite, Perseus-Pegasus à gauche),  $135 h^{-1}\text{Mpc}$  et  $150 h^{-1}\text{Mpc}$ .

Malgré tout, il est encore difficile de conclure si nous avons atteint la convergence de la vitesse reconstruite ici. Il nous faudrait pouvoir tester les résultats donnés par la reconstruction sur un volume



d'Univers plus grand. Cela nécessitera de tester sur les prochaines versions du 2MASS Redshift Survey dont la magnitude limite descendra à  $K = 11.75$  et peut-être dans un avenir proche à  $K = 12.25$ .

### Le mouvement du superamas local

Nous allons à présent étudier une autre quantité potentiellement intéressante : la vitesse du superamas local (les structures comprises dans un volume de rayon  $30 h^{-1}\text{Mpc}$ ) par rapport au fond diffus cosmologique. L'intérêt d'étudier la vitesse d'une région plus grande (telle le volume du catalogue NBG-3k) est que l'on a accès à la fois aux mesures de distances (donc la vitesse d'ensemble du superamas local) et à la fois à un volume assez large pour avoir des vitesses reconstruites vraiment fiables. Le superamas local peut servir, comme pour le Groupe Local, d'un traceur du champ de masse à grande échelle. En effet, nous nous attendons à ce que le super-amas local soit sensible à des masses à des distances plus grandes que le Groupe Local. Dans ce cas, étudier l'évolution de sa vitesse reconstruite en fonction du volume de reconstruction prend tout son sens. Nous pouvons voir l'effet de structures qui sont beaucoup plus éloignées que Shapley par exemple. On a résumé dans la table 6.6 l'évolution du bulk flow reconstruit pour le volume de  $3000 \text{ km s}^{-1}$  en utilisant deux méthodes : soit en utilisant uniquement la composante radiale et en extrapolant à l'aide de la méthode  $\chi^2$  (section 6.3.1), soit en moyennant directement les vitesses tri-dimensionnelle de tous les objets du volume (noté moyenne 3D dans la table). L'indicateur de moyenne tridimensionnel est plus fiable comme on l'a vu à la section 6.3.1. On voit en effet que le comportement de la moyenne tridimensionnelle semble plus régulier que celui donné par le  $\chi^2$  sur la figure 6.27. On voit néanmoins que nous n'arrivons à expliquer pour le moment qu'un tiers de l'amplitude du bulk flow de la bulle d'Univers de  $30 h^{-1}\text{Mpc}$  de rayon. Il peut y avoir deux explications :

- la vitesse observée peut-être faussée par des effets de projection, bien qu'une méthode POTENT (Dekel et al., 1999) appliqué sur des catalogues de distance similaires semble donner un vecteur vitesse similaire ;
- ou bien l'attracteur principal de notre super-amas local ne s'est pas encore manifesté. Cela peut être par exemple la concentration Shapley qui est assez mal résolu dans le 2MRS ( $K=11.25$ ).

Il nous semble donc que plus de données soient nécessaires pour clarifier la situation.

### Retour sur le Groupe Local - Mouvement relatif

Nous avons aussi calculé le mouvement *relatif* du Groupe Local par rapport au super-amas local, définit comme l'ensemble des structures à l'intérieur d'une sphère de  $30 h^{-1}\text{Mpc}$ . Pour cela, nous avons retranché le mouvement du superamas local (table 6.6) au mouvement du Groupe Local (haut de la table 6.5) et nous avons obtenu le bas de la table 6.5. Nous avons aussi représenté graphiquement l'amplitude et la direction de cette vitesse sur la figure 6.28. On remarque cette fois que l'on a aucun mal atteindre l'amplitude observée (on est même légèrement au-dessus) et que la direction du vecteur vitesse est quasiment dans la barre d'erreur à 95% donnée à la section 3.4.3, page 37 (22 degrés). Cela signifie que l'essentiel de nos problèmes sur le Groupe Local par rapport au CMB viennent donc bien de la dynamique à grande échelle.

## 6.4 Comparaison aux vitesses observées et paramètres cosmologiques

Une fois qu'on a un champ de vitesse reconstruit, on peut essayer de mesurer le paramètre  $\Omega_m$  par une comparaison directe avec les observations. Cependant, il va se poser un problème : comme nous ne sommes pas encore capable de reconstruire parfaitement le mouvement global du volume du catalogue NBG-3k la comparaison va être faussée. On peut néanmoins s'inspirer de la relation donnée par Juszkiewicz et al. (1990) entre la vitesse observée et reconstruite du Groupe Local :

$$\mathbf{v}_{\text{LG}} = \mathbf{v}_{\text{LG,R}} + \mathbf{v}_{\text{bulk}} - \frac{1}{3}\beta H_0 \mathbf{x}_{\text{cm}}, \quad (6.42)$$

LA SOURCE DU MOUVEMENT DU GROUPE LOCAL		
Vitesse mesurée à l'aide du dipôle du CMB : $V = 627 \pm 22 \text{ km s}^{-1}$ ; $l = 276$ , $b = 30$ .		
Référentiel	$R_{\text{rec}}$ $h^{-1}\text{Mpc}$	Vitesse reconstruite (MDC) $\text{km s}^{-1}$
CMB	10	$155$ ; $l = 143$ , $b = 25$
	20	$276$ ; $l = 236$ , $b = 31$
	30	$366$ ; $l = 252$ , $b = 37$
	40	$395$ ; $l = 271$ , $b = 28$
	50	$357$ ; $l = 249$ , $b = 18$
	60	$459$ ; $l = 225$ , $b = -13$
	80	$465$ ; $l = 215$ , $b = 37$
	100	$516$ ; $l = 201$ , $b = 28$
	120	$602$ ; $l = 211$ , $b = 27$
	120 (S)	$651$ ; $l = 210$ , $b = 37$
	150	$611$ ; $l = 249$ , $b = -20$
	150 (S)	$475$ ; $l = 205$ , $b = 43$

Vitesse mesurée du Groupe Local dans le référentiel du super-amas local : $V = 302 \pm 22 \text{ km s}^{-1}$ ; $l = 241$ , $b = 37$ .		
30 $h^{-1}\text{Mpc}$	30	$266$ ; $l = 244$ , $b = 32$
	40	$225$ ; $l = 261$ , $b = 31$
	50	$247$ ; $l = 231$ , $b = 25$
	60	$353$ ; $l = 215$ , $b = 0.29$
	80	$356$ ; $l = 208$ , $b = 28$
	100	$371$ ; $l = 204$ , $b = 21$
	120	$472$ ; $l = 207$ , $b = 21$
	120 (S)	$463$ ; $l = 207$ , $b = 26$
	150	$426$ ; $l = 230$ , $b = -12$
	150 (S)	$343$ ; $l = 187$ , $b = 36$

TAB. 6.5: *Origine du mouvement du Groupe Local par rapport au CMB* – Cette table résume les mesures du mouvement reconstruit du Groupe Local en fonction de la profondeur de reconstruction. Deux profondeurs de reconstruction sont annotés d’un “(S)” : cela indique que l’on a utilisé la correction d’incomplétude indiqué page 120 en fixant un seuil maximum de correction par amplification de la luminosité à 2 (ce qui correspond à une distance de  $100 h^{-1}\text{Mpc}$ ).

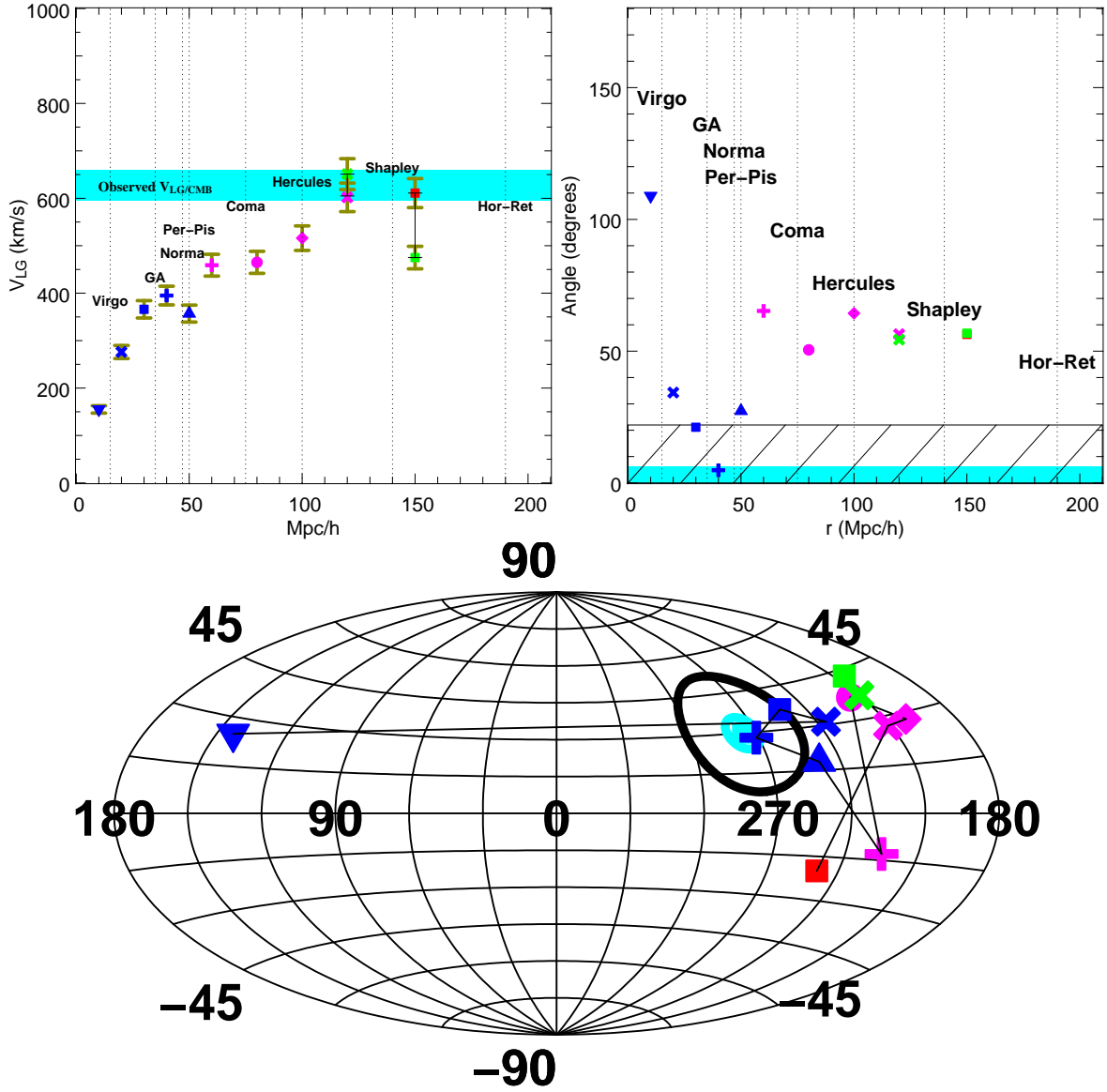


FIG. 6.25: *Origine du mouvement du Groupe Local par rapport au CMB* – Cette figure résume les résultats de la table 6.5 d’une manière graphique. Sur le panneau en haut à gauche, on a représenté l’amplitude de la vitesse reconstruite du Groupe Local en fonction de la profondeur de reconstruction. Sur le panneau en haut à droite, on a représenté l’écart angulaire entre la vitesse reconstruite du Groupe Local et celle observée grâce au CMB. Les zones en bleu cyan sur les trois figures indiquent la barre d’erreur observationnelle. Les barres d’erreur colorées en beige dénote l’erreur minimale auquel on s’attend de par l’hypothèse sur  $\Omega_m$  qui a été prise ( $\sim 5\%$ ). Sur le panneau du bas, on a représenté la direction de la vitesse reconstruite en fonction de la profondeur de reconstruction. Les symboles et couleurs sont les mêmes sur les trois panneaux. Le symbole carré cyan correspond au vecteur vitesse observé du Groupe Local par rapport au CMB. Nous avons ajouté sur les figures du haut les distances de structures majeures dans l’Univers local (Per-Pis est le superamas de Perseus-Pisces et Hor-Ret le superamas de Horologium-Reticulum). Les symboles verts correspondent aux deux reconstructions “(S)” de la table 6.5. On a entouré la région du ciel située à moins de 20 degrés de la direction donnée par le CMB par un contour noir épais sur le panneau du bas. Ce même contour correspond à la zone hachurée dans le panneau en haut à droite.

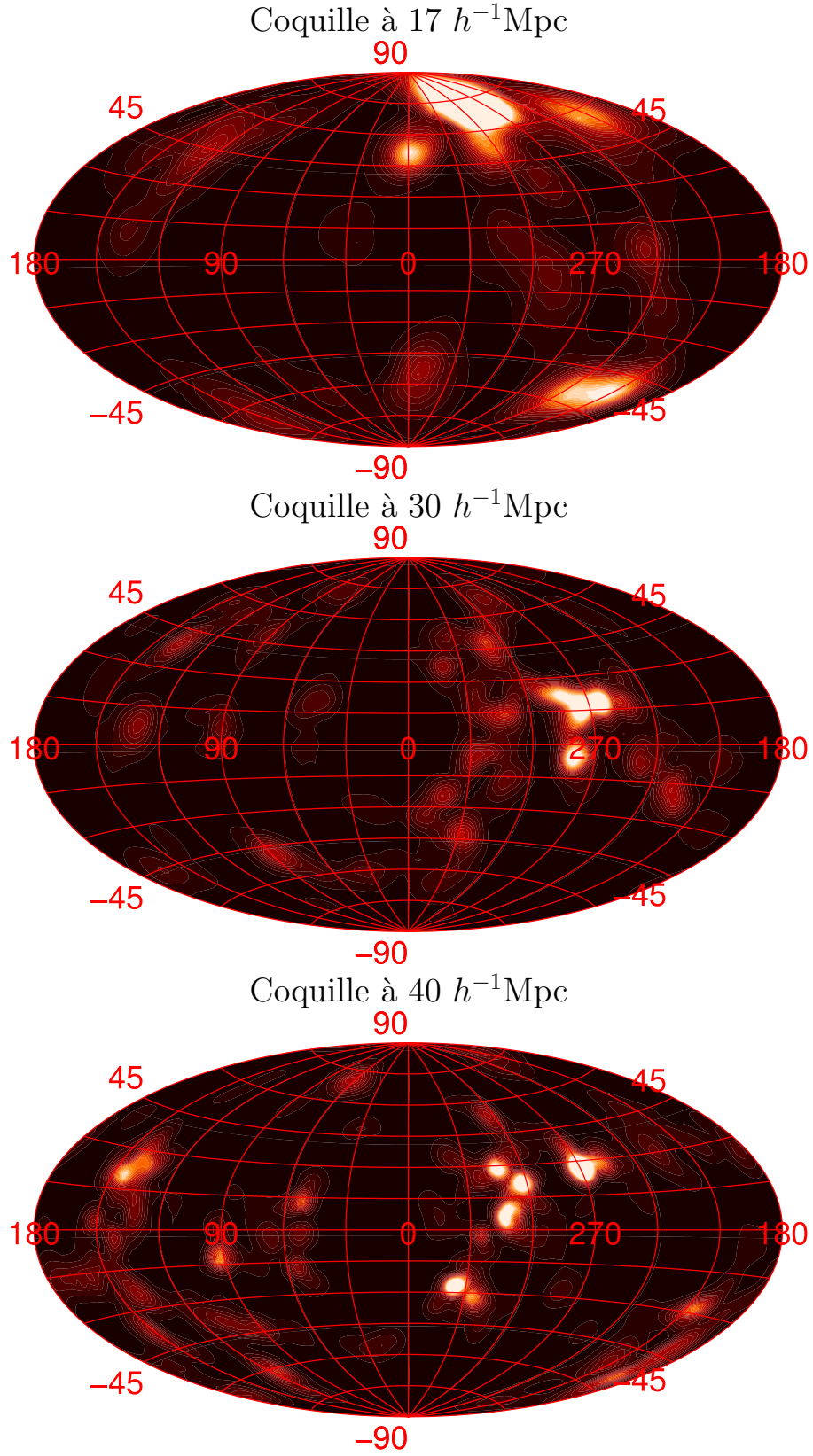


FIG. 6.26: *Illustration du champ de densité à différentes distances de l'observateur* – On a représenté ici différentes coquilles de rayon  $R$  indiqué au-dessus de chacun des panneaux, extraites du champ de densité issu catalogue 2MRS par filtrage adaptatif. Le codage de couleur est linéaire avec la masse. On reconnaît les structures majeures indiquées dans la figure 2.3 et la table 2.1.

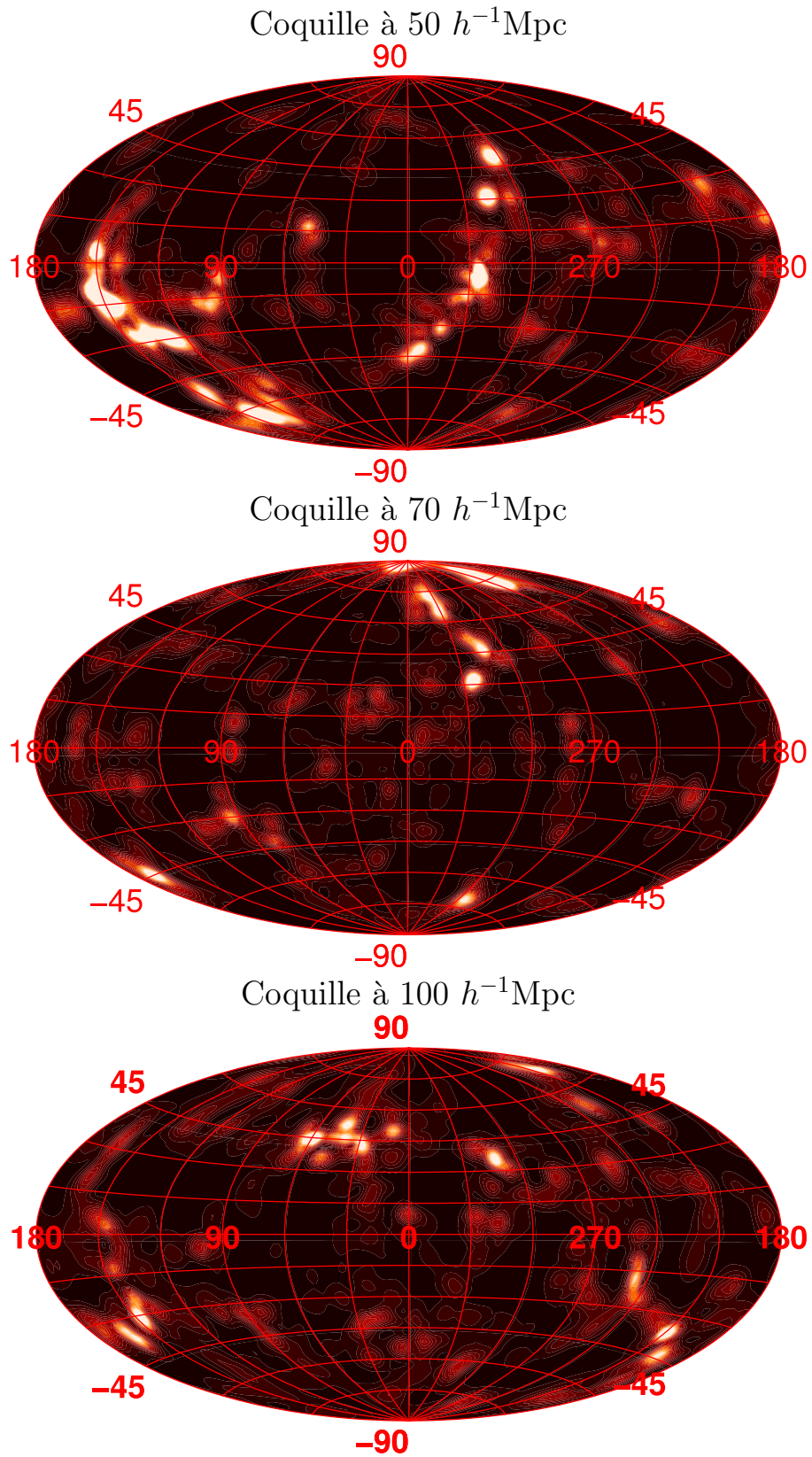


FIG. 6.26 – Suite.

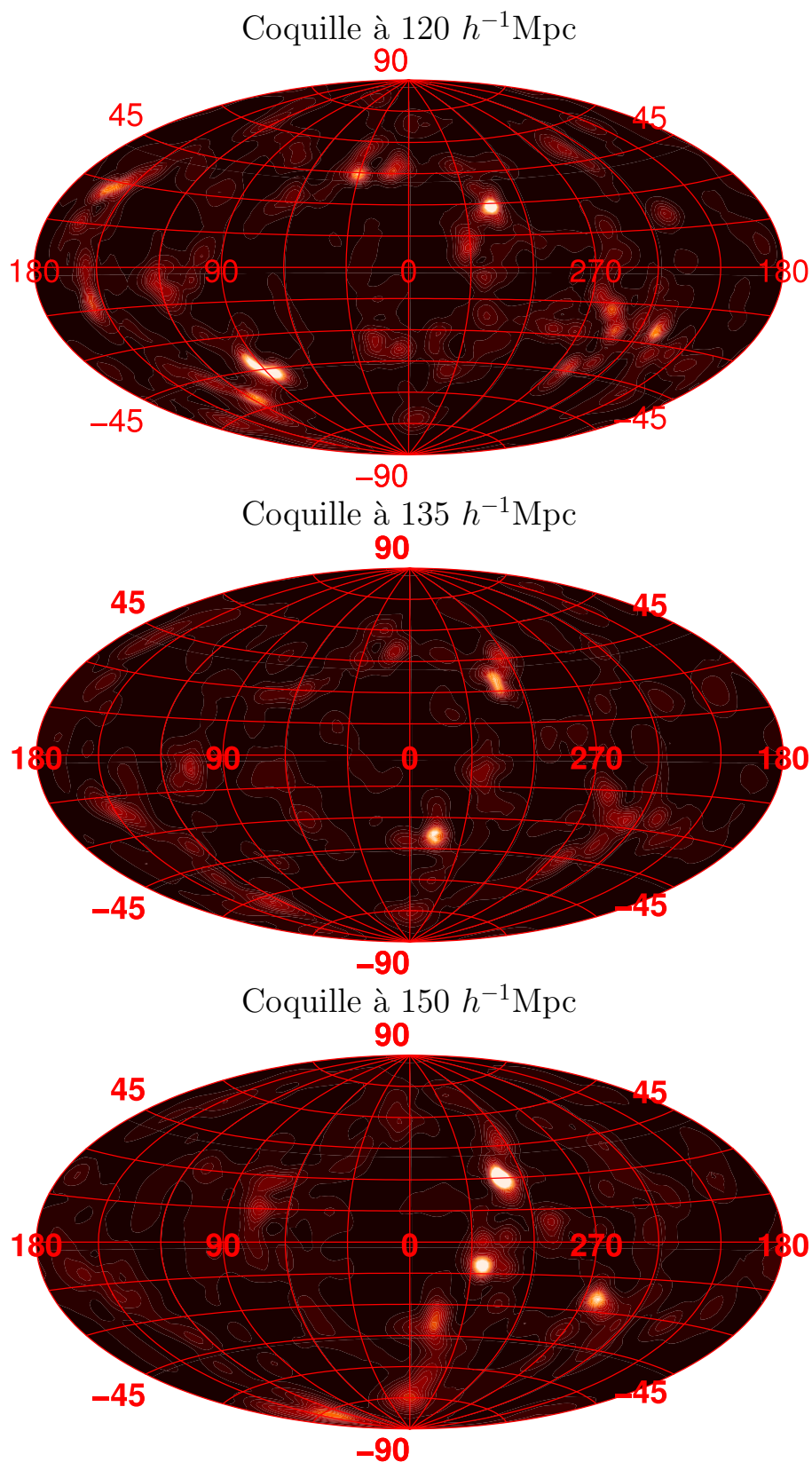


FIG. 6.26 – Suite.

LA SOURCE DU MOUVEMENT D'ENSEMBLE DU VOLUME DE RAYON $30 h^{-1}\text{Mpc}$ .		
Vitesse mesurée grâce aux vitesses observées dans le référentiel du CMB : $V = 398 \pm 22 \text{ km s}^{-1}$ ; $l = 298 \pm 4$ , $b = 19 \pm 2$ .		
$R_{\text{rec}}$ $h^{-1}\text{Mpc}$	Méthode	Vitesse reconstruite (MDC) $\text{km s}^{-1}$
30	$\chi^2$ moyenne 3D	$49 \pm 24$ ; $l = 2 \pm 14$ , $b = 8 \pm 13$ $110$ ; $l = 278$ , $b = 44$
40	$\chi^2$ moyenne 3D	$98$ ; $l = 323 \pm 5$ , $b = -2 \pm 4$ $176$ ; $l = 283$ , $b = 23$
50	$\chi^2$ moyenne 3D	$102$ ; $l = 326$ , $b = -26$ $145$ ; $l = 278$ , $b = 2$
60	$\chi^2$ moyenne 3D	$154$ ; $l = 348$ , $b = -71$ $157$ ; $l = 257$ , $b = -42$
80	$\chi^2$ moyenne 3D	$105$ ; $l = 33$ , $b = 29$ $130$ ; $l = 247$ , $b = 56$
100	$\chi^2$ moyenne 3D	$102$ ; $l = 87$ , $b = 29$ $154$ ; $l = 194$ , $b = 44$
120	$\chi^2$ moyenne 3D	$92$ ; $l = 43$ , $b = 31$ $144$ ; $l = 226$ , $b = 46$
120 (S)	$\chi^2$ moyenne 3D	$162$ ; $l = 54$ , $b = 36$ $213$ ; $l = 219$ , $b = 60$
150	$\chi^2$ moyenne 3D	$272$ ; $l = 342$ , $b = -42$ $255$ ; $l = 286$ , $b = -28$
150 (S)	$\chi^2$ moyenne 3D	$142$ ; $l = 33$ , $b = 2$ $169$ ; $l = 250$ , $b = 46$

TAB. 6.6: *Origine du mouvement du superamas local (volume de rayon  $30 h^{-1}\text{Mpc}$ )* – Cette table résume les mesures du mouvement reconstruit du superamas local en fonction de la profondeur de reconstruction. Deux profondeurs de reconstruction sont annotés d’un “(S)” : cela indique que l’on a utilisé la correction d’incomplétude indiqué page 120 en fixant un seuil maximum de correction par amplification de la luminosité à 2 (ce qui correspond à une distance de  $100 h^{-1}\text{Mpc}$ ).

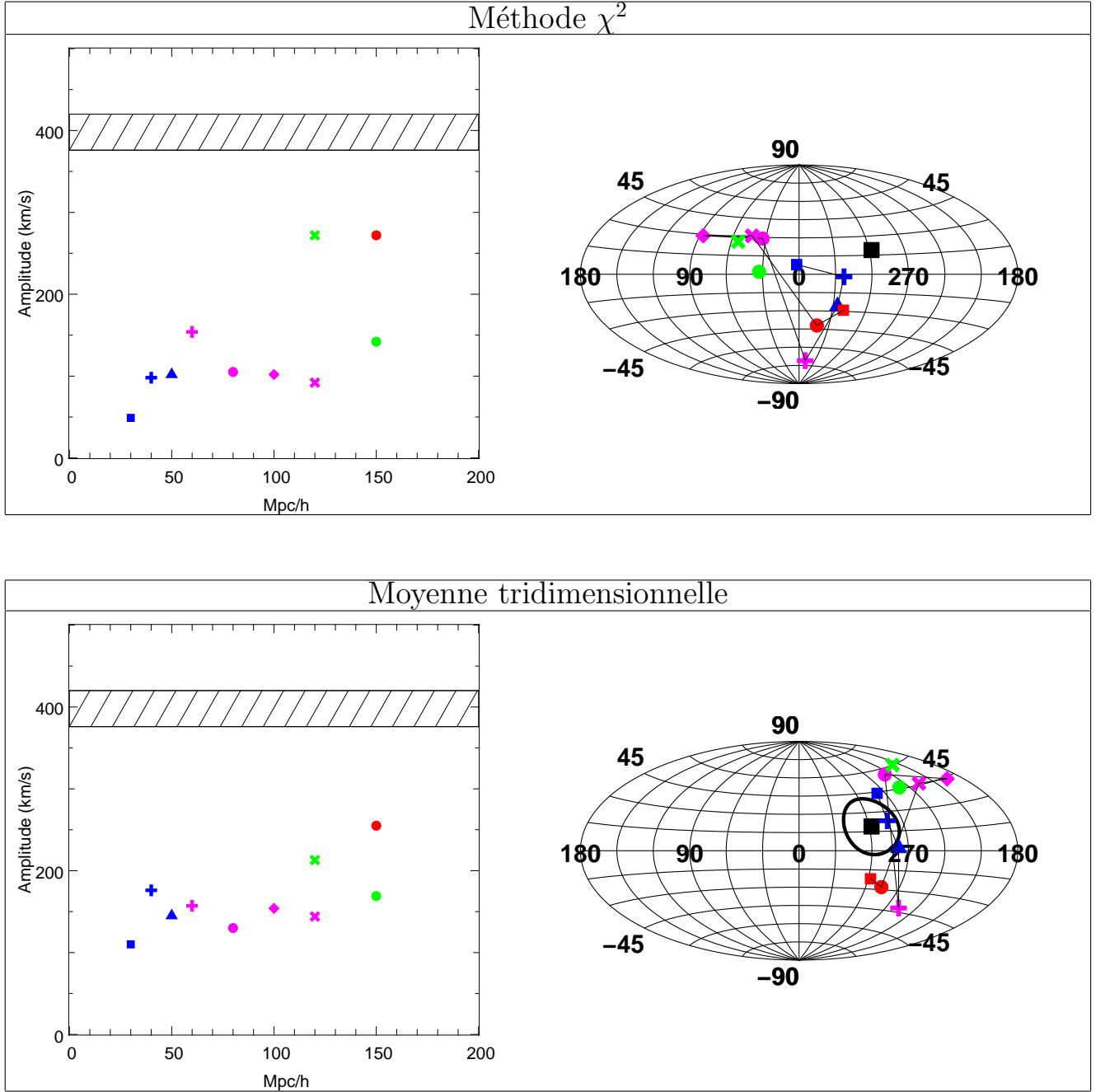


FIG. 6.27: *Origine du mouvement du superamas local (volume de rayon  $30 h^{-1} \text{Mpc}$ )* – Cette figure résume les résultats présentés dans la table 6.6 d’une manière graphique. En haut, on a résumé dans les panneaux du haut les résultats donnés par la méthode  $\chi^2$ . Dans les panneaux du bas, on a les résultats obtenus par la méthode de la moyenne tridimensionnelle des vitesses des objets du volume de rayon  $3000 \text{ km s}^{-1}$ . Sur les panneaux de gauche, on a représenté l’amplitude du vecteur vitesse alors que les panneaux de droite donnent la direction du vecteur vitesse en projection aitoft. Les symboles et couleurs utilisées sur les graphiques représentent les mêmes volumes de reconstruction. Les symboles verts correspondent aux deux reconstructions “(S)” de la table 6.6.



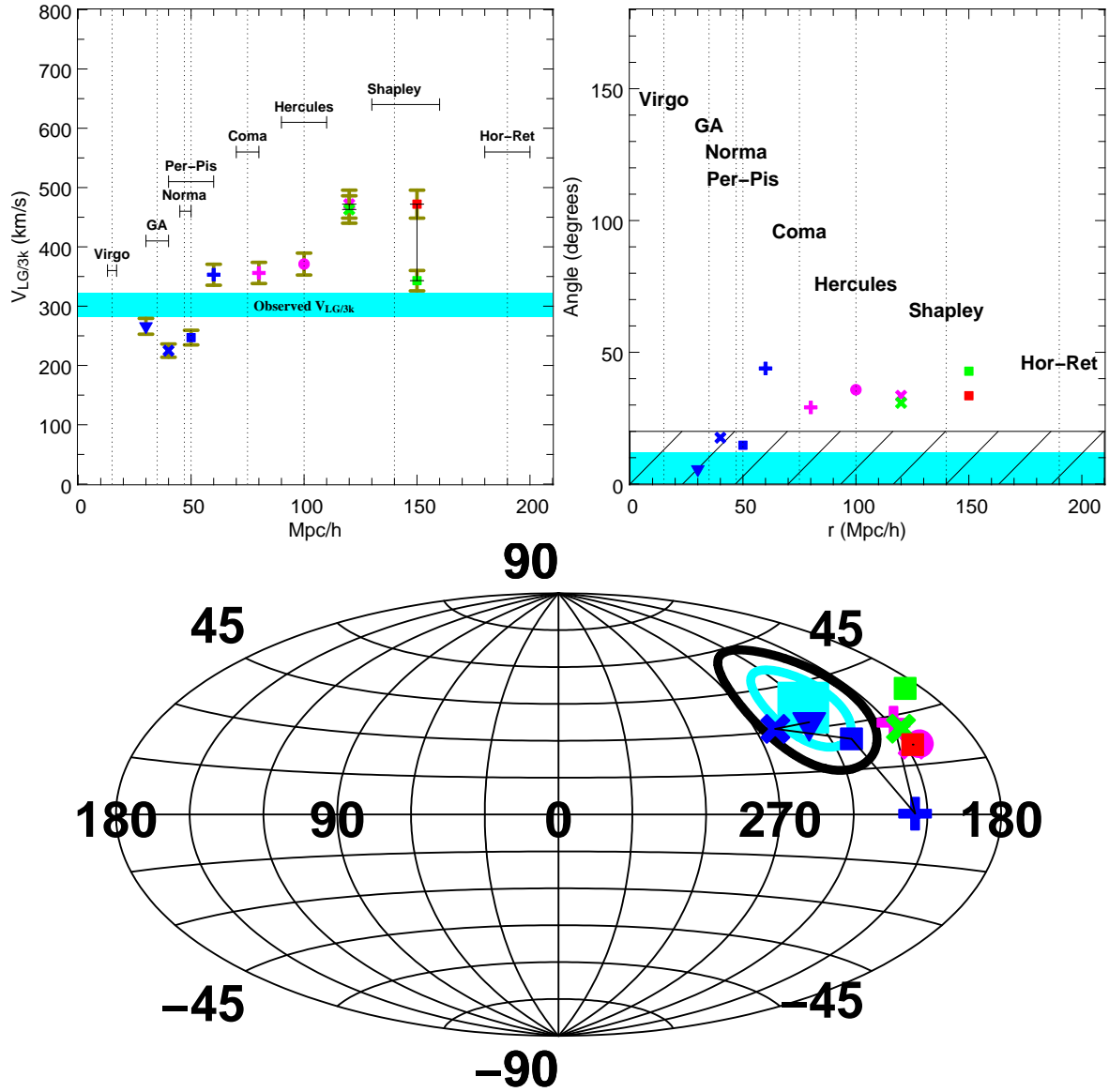


FIG. 6.28: *Origine du mouvement du Groupe Local par rapport au super-amas local* – Même chose que pour la figure 6.25 mais pour le mouvement pris par rapport au super-amas local (volume sphérique de rayon  $3000 \text{ km s}^{-1}$ ).

avec  $\mathbf{v}_{\text{LG}}$  la vitesse du Groupe Local par rapport au CMB,  $\mathbf{v}_{\text{LG,R}}$  la vitesse calculée du Groupe Local à partir d'un relevé de galaxie de profondeur  $R$ ,  $\mathbf{v}_{\text{bulk}}$  le mouvement d'ensemble de ce relevé,  $\mathbf{x}_{\text{cm}}$  la position du centre de masse de ce relevé au sens

$$\mathbf{x}_{\text{cm}} = \int_{\text{sphère de rayon } R} d^3\mathbf{r} \mathbf{r} \delta(\mathbf{r}). \quad (6.43)$$

Cette relation, dérivée de la théorie linéaire, nous donne une intuition que dans le cas de la reconstruction il faudra aussi faire attention de comparer les bonnes vitesses. Pourvu que l'on ne s'approche pas trop des bords, cette relation doit encore être valide. Nous comparerons donc les vitesses *observées* retranchées du mouvement d'ensemble *observé* aux vitesses *reconstruites* retranchées du mouvement d'ensemble *reconstruit*. Comme nous devons comparer des champs de vitesse radiale, nous utiliserons les vitesses d'ensemble estimées par la méthode  $\chi^2$ .

#### 6.4.1 Analyse des vitesses simulées et reconstruites d'une simulation $\Lambda$ CDM

Dans la figure 6.29, on montre le résultat de la comparaison entre le champ de vitesse radial, obtenu par filtrage adaptatif, observé et reconstruit avant et après soustraction du bulk flow mesuré dans la table 6.2. Il apparaît en effet que le scatter soit plus corrélés après la soustraction, par contre la pente ne semble pas changer d'un panneau à l'autre. Par contre, le  $\Omega_m$  mesuré n'est pas grandement amélioré dans ce cas. L'effet de la soustraction, serait donc d'augmenter la corrélation sans toutefois changer la pente générale de la PDF. On s'attend à un tel effet car, si on écrit la relation entre champ de vitesse reconstruit et observé

$$v_{\text{rec},r}(\mathbf{r}) = v_{\text{obs},r}(\mathbf{r}) + (\mathbf{v}_{\text{rec,bulk}} - \mathbf{v}_{\text{obs,bulk}}) \cdot \hat{\mathbf{r}} = v_{\text{obs},r}(\mathbf{r}) + \delta\mathbf{v} \cdot \hat{\mathbf{r}} \quad (6.44)$$

avec  $\hat{\mathbf{r}}$  le vecteur unitaire  $\hat{\mathbf{r}} = \mathbf{r}/\|\mathbf{r}\|$ . On va couper l'espace en deux parties séparées par le plan  $\mathcal{P}_0$  définit par l'équation  $\delta\mathbf{v} \cdot \hat{\mathbf{r}} = 0$ . On pose  $\mathcal{P}_+$  (respectivement  $\mathcal{P}_-$ ) la région de l'espace où  $\delta\mathbf{v} \cdot \hat{\mathbf{r}}$  est positif (respectivement négatif). Tous les objets du catalogues dont la position est dans  $\mathcal{P}_+$  auront une vitesse radiale reconstruite augmentée par rapport à la vitesse observée, avec un décalage maximum de  $\|\delta\mathbf{v}\|$ . On observe le phénomène opposé dans la partie  $\mathcal{P}_-$  : les objets ont une vitesse radiale reconstruite systématiquement diminuée par rapport à la vitesse observée, avec le même décalage maximum. On se retrouve donc avec, grosso-modo, une distribution de probabilité jointe ( $v_{r,\text{rec}}, v_{r,\text{obs}}$ ) bimodale, avec deux parties bien corrélées au champ de vitesse mais avec un décalage systématique par rapport à la diagonale du type illustré sur la figure 6.30. Au contraire les objets proche du plan  $\mathcal{P}_0$  ne sont quasiment pas affectés. Cela peut conduire à une mauvaise estimation de  $\Omega_m$  si on n'effectue pas la correction de bulk flow. Bien sur, cet effet peut être amplifié en présence de la ZOA. En effet si celle-ci est confondue avec  $\mathcal{P}_0$ , il nous manquera la partie non-affectée de la densité de probabilité, ce qui renforce l'effet de séparation des deux régions  $\mathcal{P}_+$  et  $\mathcal{P}_-$ .

Une fois corrigé de cet effet, on trouve une forme de PDF plus homogène (figure 6.29). Néanmoins, la mesure de la pente du nuage conduit à  $\Omega_m = 0.195$  par la méthode des moments dans un volume de  $3000 \text{ km s}^{-1}$ , au lieu de  $\Omega_{\text{eff}} = 0.279$ . Cette valeur est par contre en parfait accord avec la densité moyenne dans une sphère de  $20 h^{-1} \text{ Mpc}$  (nous avons par ailleurs vérifié que la même mesure prise dans  $20 h^{-1} \text{ Mpc}$  donne toujours  $\Omega_m = 0.195$ ). On est donc face à un problème majeur de variance cosmique où on traite un catalogue profondément sous-dense proche de l'observateur et qui ne converge vers la densité moyenne que lentement (en effet à  $100 h^{-1} \text{ Mpc}$  on a une fluctuation négative  $\delta_{100} = -0.028$ ). Ce catalogue virtuel est probablement un peu extrême et il serait nécessaire d'avoir une statistique sur un grand nombre de catalogues virtuels afin de pouvoir conclure sur la variance sur la mesure de  $\Omega_m$  due à la méthode.

#### 6.4.2 Vitesses observées et reconstruites dans le catalogue 2MRS/NBG-3k

On va chercher ici à faire une comparaison entre la composante suivant la ligne de visée du champ de vitesse reconstruit et observé. Pour calculer le champ de vitesse à l'aide du filtre adaptatif, on a

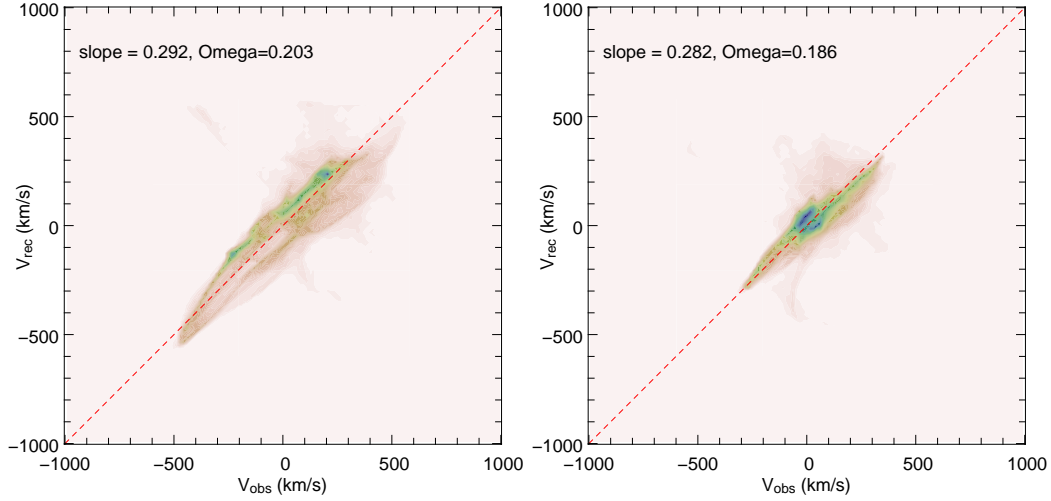


FIG. 6.29: *Comparaison vitesses reconstruites/simulées dans le Mock2MRS-CMB* – Les deux panneaux donnent la distribution de probabilité jointe d’avoir un couple  $(v_{r,\text{rec}}, v_{r,\text{sim}})$ , avec  $v_{r,\text{rec}}$  et  $v_{r,\text{sim}}$  étant respectivement les composantes selon la ligne de visée des vitesses reconstruites et simulées. On a indiqué quel  $\Omega_m$  est à la fois mesuré (par comparaison avec la vitesses simulées) et supposé pour l’affichage du graphe. Le panneau de gauche donne le résultat de la comparaison si on n’effectue aucune correction. Le panneau de droite donne ce même résultat mais après soustraction à chacun des champs des bulk flow mesurés à l’aide des champs non corrigés tel qu’expliquer dans le texte.

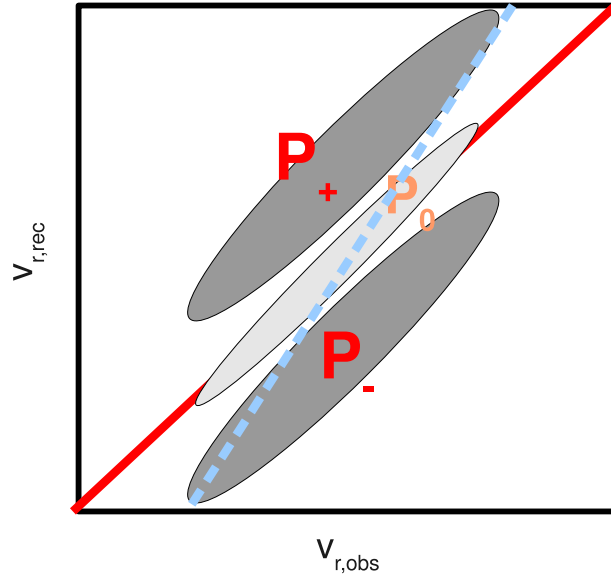


FIG. 6.30: *Illustration de l’effet du bulk flow sur la comparaison vitesse-vitesse* – On a représenté graphiquement ici l’argument donné dans le texte principal (section 6.4.1). On sépare les points en trois sous-ensemble  $\mathcal{P}_+$ ,  $\mathcal{P}_0$  et  $\mathcal{P}_-$ . Chacun de ces ensemble a une bonne pente, mais peut produire une mauvaise estimation de  $\Omega_m$ . Cet effet est particulièrement visible dans les observations (figure 6.31)

utilisé le sous-ensemble de traceur du champ de vitesse ayant été à la fois reconstruit et ayant eu une distance mesurée, tout en étant dans une sphère de rayon  $30 h^{-1}\text{Mpc}$ . Le résultat de la comparaison entre les deux champs est donné figure 6.31. On constate que les deux panneaux semblent avoir la même pente (voir par exemple les deux queues de distribution parallèle l’une par rapport à l’autre dans les panneaux de gauche). Néanmoins, après soustraction des bulk flow on voit clairement apparaître que l’on décrit bien le même champ de vitesse : la corrélation devient évidente. On a pour le moment appliqué uniquement la méthode des moments à cette distribution. Celle-ci indique que le  $\Omega_m$  serait préférentiellement de l’ordre de 0.28. Les barres d’erreur données par cette méthode sont suffisamment grandes néanmoins pour inclure tous les  $\Omega_m$  entre 0 et 1 bien que graphiquement on puisse constater que ce n’est pas le cas. Il reste donc à trouver une analyse en fonction de vraisemblance qui soit à la fois robuste aux queues de distribution (comme on l’a développé à la section 4.3) et n’étant pas trop soumis aux priors. Cette comparaison finale montre que l’on reconstruit bien le bon champ de vitesse dans l’Univers local pourvu qu’on s’affranchisse de l’effet de bulk flow.

## 6.5 Conclusion

Nous avons vu dans ce chapitre comment utiliser une méthode de reconstruction lagrangienne des trajectoires de galaxies de notre Univers local (section 6.1.2 pour NBG-3k, 6.2.3 et 6.2.4 pour 2MRS). Nous avons pu ensuite utiliser ces trajectoires pour comprendre quelle était la source du mouvement du Groupe Local (section 6.3.4). Il apparaît que la “construction” de notre vitesse est dans la moyenne pour un univers  $\Lambda\text{CDM}$  d’après la théorie développée par Juszkiewicz et al. (1990) à partir de la linéaire de croissance des fluctuations. En effet on s’attend à ce que, pour  $\Omega_m = 0.30$  et  $\Omega_\Lambda = 0.70$ , en moyenne on ait convergé à 93% de la valeur finale de notre vecteur vitesse pour un relevé de  $100 h^{-1}\text{Mpc}$  de rayon. Cette valeur peut sembler différente de ce qu’on a trouvé ici ( $\sim 74\%$ ) mais la distribution autour de 93% est assez large pour que notre valeur soit acceptable. Nous devons néanmoins nuancer ce propos. En effet, nous n’avons pas encore vu de stricte convergence du mouvement du Groupe Local : l’amplitude arrive juste à la valeur observée grâce au fond diffus cosmologique mais la direction est encore éloignée d’environ 50 degrés. De plus, il se peut que nous sous-estimions l’impact de structures massives plus éloignées. Le super-amas de Shapley n’est par exemple pas bien résolu dans le 2MRS  $K=11.25$ . L’impact du super-amas Horologium-Reticulum (voire Pisces-Cetus), à  $\sim 160 h^{-1}\text{Mpc}$ , n’a pas été pris en compte. Nous sommes néanmoins rassuré de voir que notre vitesse *relative* par rapport au super-amas local semble bien proche de la convergence. Donc l’essentiel de nos problèmes viennent des régions d’au-delà du Grand Attracteur (Hydra-Centaurus). Il nous faut souligner le fait que la méthode a été entièrement validée par l’usage des simulations numériques et des catalogues virtuels.

Un deuxième point important de ce chapitre est la première reconstruction du champ de vitesse de l’Univers local par méthode lagrangienne sur un volume conséquent. La composante selon la ligne de visée du champ de vitesse reconstruit a été ensuite comparée à celle observée avec succès. Il reste néanmoins à affiner l’approche en fonction de vraisemblance afin d’obtenir un résultat quantitativement bien défini sur les paramètres cosmologiques. Une première estimation montre que  $\Omega_m \simeq 0.28$  dans notre Univers local, ce qui est en accord avec la valeur estimée à partir de l’expérience WMAP (Dunkley et al., 2008). Encore une fois, la méthode utilisée pour obtenir ce résultat a été validée grâce aux simulations numériques.

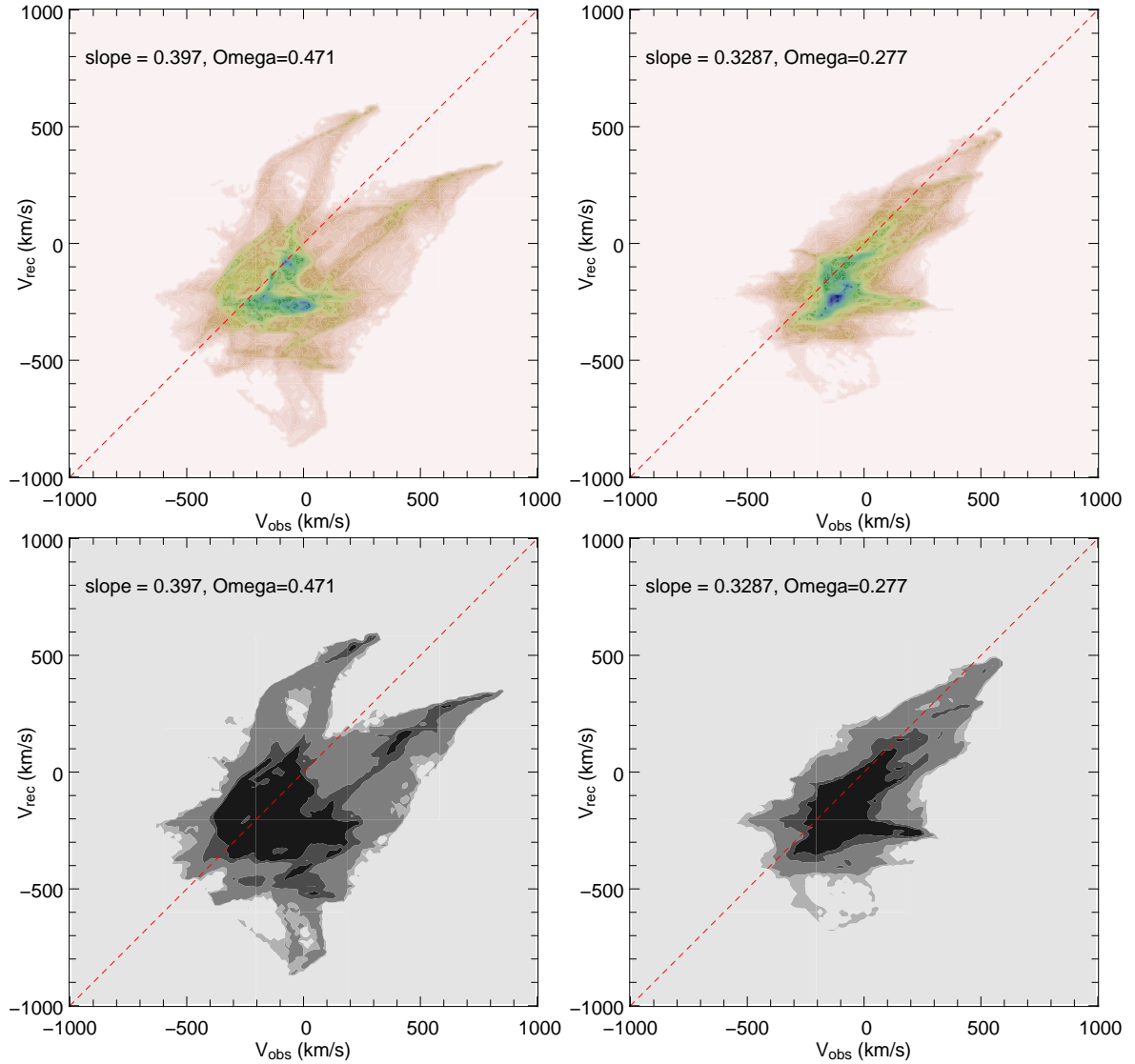


FIG. 6.31: *Comparaison des vitesses reconstruites avec 2MASS aux vitesses observées dans une sphère de rayon  $30 h^{-1} \text{Mpc}$*  – Afin d’obtenir les vitesses observées, on a supposé ici  $H_0 = 80 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Ce choix n’affecte le diagramme que par une translation selon l’axe des X de la distribution de probabilité des variables  $(V_{\text{rec}}, V_{\text{obs}})$ . On a représenté sur les panneaux de gauche la distribution que l’on obtient si on ne retranche pas les bulk flow (méthode  $\chi^2$ ) respectifs des vitesses reconstruites et observées. Sur les panneaux de droite, ces bulk flow ont été soustraits. Les panneaux du bas donnent la même chose que les panneaux du haut mais tracent uniquement les contours à 50%, 68%, 95% et 99% (du noir au gris clair). On a indiqué la pente et le  $\Omega_m$  mesuré correspondant avec et sans soustraction des bulk flow.

## 7 Conclusion et perspectives

AU DÉBUT DE CETTE THÈSE, nous nous étions fixés comme but d’étudier le comportement et les limites des reconstructions lagrangiennes pour l’utilisation sur des catalogues de galaxie, afin d’appliquer ces méthodes sur le catalogue 2MASS Redshift Survey en vue d’obtenir un champ de vitesse beaucoup plus précis qu’en utilisant la théorie linéaire. Nous avons d’abord argumenté de l’intérêt des champs de vitesse (section 2.4.1, page 18). Bien que très difficile à mesurer, ils recèlent énormément d’informations sur la dynamique à grande échelle (le spectre de puissance des fluctuations du champ de vitesse a plus d’amplitude aux grandes échelles que celui des fluctuations de densité) et à plus petite échelle (le champ de vitesse est un traceur quasi direct de la masse dans les amas). Nous avons ensuite détaillé quelques méthodes de reconstruction qui permettent d’accéder à cette information sans nécessairement avoir accès au vrai champ de vitesse dans l’Univers. L’intérêt résidant alors à effectuer une comparaison a posteriori entre les vitesses reconstruites et observées. Parmi ces méthodes, nous nous sommes par la suite focalisés sur la reconstruction de Monge-Ampère-Kantorovitch (MAK, section 2.4.6) qui offre une solution unique au problème de reconstruction des orbites des galaxies. C’est un point important car on souhaite appliquer cette méthode à de grands catalogues (en volume et en nombre de galaxies) et on ne peut se permettre dans ce cas de séparer les solutions probables de celles qui ne le sont pas. La recherche de cette solution a été facilitée par l’utilisation d’un algorithme performant (annexe B), dont le code a été parallélisé et exécuté sur des supercalculateurs, par exemple au Centre de calcul du CEA sur le calculateur Platine. Nous avons ensuite étudié la qualité des solutions reconstruites par la méthode MAK à l’aide d’une simulation à  $N$ -corps (chapitre 3, page 29) : du point de vue des orbites (section 3.3, page 31), du champ de vitesse (section 3.4, page 35) et de la reconstruction des fluctuations initiales de la simulation (section 3.5, page 40). Cette étude a montré que MAK est en théorie capable de reconstruire le champ de vitesse jusqu’à une échelle d’environ  $5 h^{-1}\text{Mpc}$  et le champ de densité primordial à environ  $12 h^{-1}\text{Mpc}$ . Une des premières applications possibles est de mesurer  $\Omega_m$ . Nous avons construit pour cela une analyse bayésienne au chapitre 4. Nous avons exploité ensuite ces outils statistiques pour mesurer  $\Omega_m$  et les biais de reconstruction du champ de vitesse sur des catalogues virtuels de galaxies (chapitre 5). Nous avons trouvé que dans la majeure partie des cas la reconstruction est à même de reproduire le champ de vitesse, mais qu’il existe aussi des biais fondamentaux qu’il est difficile de corriger (par exemple la masse manquante et le rapport  $M/L$ ). Aussi, nous avons trouvé que l’approche bayésienne “naïve” est assez sensible aux a priori mis dans la fonction de vraisemblance. Il faudra manier avec prudence le résultat donné par celle-ci (en comparant par exemple avec la méthode des moments). Nous pensons notamment qu’il nous faudrait tenir compte de la corrélation du champ de vitesse dans la fonction de vraisemblance (qui est malheureusement difficile à implémenter numériquement).

NOUS NOUS SOMMES ENSUITE TOURNÉS vers la reconstruction du champ de vitesse réel de notre Univers à l’aide de deux catalogues : le NBG-3k (qui fournit les distances) et le 2MRS (qui fournit les redshifts sur tout le ciel et sur un volume plus conséquent). Nous avons entrepris de préparer le catalogue 2MRS pour une reconstruction. Il a fallu pour cela

- lier les galaxies du 2MRS à celles du NBG-3k,
- identifier les groupes de galaxies et comprimer les “finger-of-god”,
- estimer le rapport  $M/L$  des galaxies du 2MRS,
- corriger l’incomplétude,
- et corriger l’obscurcissement du plan galactique.

Ce catalogue nous a permis d’abord d’étudier quelle est la source du mouvement du Groupe Local par rapport au CMB à l’aide d’une série de reconstructions de volume différents. Pour la première fois,

il est possible d'étudier la dynamique *non-linéaire* donnant lieu à notre mouvement. Il semble que nous atteignons juste la convergence de l'amplitude mais pas encore de la direction à environ  $150 h^{-1}\text{Mpc}$ . Cela est probablement en partie dû à l'incomplétude du catalogue 2MRS qui est substantielle à ces distances. Cela pourra être amélioré à l'aide des versions suivantes du catalogue où la magnitude apparente limite sera plus élevée. Le catalogue 2MRS nous a aussi permis d'étudier pour la première fois la source du mouvement du superamas local. Celui-ci étant plus volumineux que le Groupe Local, il est sensible à des masses plus éloignées. Nous avons ainsi constaté qu'on était encore éloigné de la convergence vers les observations de l'amplitude de la vitesse. Enfin, il a été aussi possible d'avoir une première estimation grossière de  $\Omega_m$  à l'aide de la comparaison entre vitesses observées et reconstruites. Cette comparaison montre que les deux champs de vitesse sont fortement corrélés entre eux (tout au moins visuellement). Nous avons pu ainsi trouver un  $\Omega_m \simeq 0.28$ , qui est en accord avec les observations du fond diffus cosmologique. Il reste néanmoins à faire une estimation quantitative précise à l'aide d'une analyse en fonction de vraisemblance qui s'est avérée plus sensible aux a priori théoriques que ce qu'on ne l'imaginait au départ. Une piste d'amélioration a été proposée à la section 4.3.4 (page 63).

**N**OUS AIMERIONS CONCLURE sur quelques perspectives de la méthode de reconstruction MAK qui peuvent se décliner en trois grandes parties : exploitation des données de reconstruction, amélioration de la méthode, exploration de la dynamique de la matière aux très petites échelles.

### Autres applications des méthodes de reconstruction

A partir des déplacements reconstruits, il est possible de tirer un grand nombre d'informations sur le spectre de puissance sur une nouvelle gamme d'échelles et d'étudier les paramètres cosmologiques autres que  $\Omega_m$ . Il est aussi possible d'effectuer de cette manière une comparaison entre les propriétés statistiques de notre Univers local et de celui dont le fond diffus cosmologique nous donne une image. Il se peut qu'il y ait des différences car il s'agit de deux réalisations indépendantes du processus stochastique sous-jacent à la formation des structures à grande échelle. Le gros avantage de la reconstruction d'orbite est qu'elle nous permet aussi d'accéder aux conditions initiales notre Univers local. On pourrait ainsi réaliser des simulations contraintes d'une qualité sans précédent. Ces simulations nous permettraient en retour de mieux comprendre le rapport  $M/L$  et, d'ainsi, améliorer la reconstruction.

### Extension de la reconstruction MAK dans le régime Euler-Poisson

Un point que nous avons laissé de côté dans cet exposé est l'extension possible de la théorie de la reconstruction MAK dans des régimes plus non-linéaires. Si nous reprenons l'action d'Euler-Poisson, exprimée dans un contexte cosmologique, avec le terme cinétique de Nusser & Branchini (2000) :

$$S[\boldsymbol{\theta}(\mathbf{x}, D), \rho(\mathbf{x}, D)] = \int_0^{D_{\text{final}}} dD \int_{\mathbf{x}} d^3\mathbf{x} \rho(\mathbf{x}, D) \left[ \frac{1}{2} D^{3/2} \|\boldsymbol{\theta}(\mathbf{x}, D)\|^2 - \frac{3\Omega_m}{2D^{1/2} f(\Omega_m)^2} \phi_g(\mathbf{x}, D) \right] \quad (7.1)$$

avec  $D$  le facteur de croissance des fluctuations linéaires pris comme axe des temps,  $\rho(\mathbf{x}, D)$  la densité eulérienne de matière à l'instant  $D$  et à la position  $\mathbf{x}$ ,  $\boldsymbol{\theta} = d\mathbf{x}/dD$  le champ de vitesse comobile et

$$\text{div}_{\mathbf{x}} \mathbf{g} = \delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \quad (7.2)$$

le champ de gravité dû aux fluctuations de matière. Il a été montré dans l'annexe D de Brenier et al. (2003) qu'une action extrêmement voisine de celle-ci (à une constante près) est convexe, et n'admet qu'une seule et unique solution. Il n'existe donc qu'un minimum à l'action (7.1). Nous nous proposons de le trouver par une manière itérative de réduction de la valeur de l'action. Pour les besoins de l'algorithme, nous allons à présent décrire nos champs à l'aide de particules traceuses de masses égales dont la valeur est  $m$ , de positions  $\mathbf{x}_i$  et de vitesses  $\boldsymbol{\theta}_i$ , dont on suppose avoir déterminé des

trajectoires *ansatz*. On souhaite trouver les trajectoires de ces particules qui satisfont avec une meilleure approximation les équations d'Euler-Poisson. Avec une description particulière, l'action discrétisée devient alors :

$$S[\mathbf{x}_i(D)] = m \sum_{i=1}^N \int_0^{D_{\text{final}}} D^{3/2} dD \left[ \frac{1}{2} \|\boldsymbol{\theta}_i\|^2 - \kappa(D) \phi_g(\mathbf{x}_i, D) \right] \quad (7.3)$$

avec

$$\kappa(D) = \frac{3\Omega_m}{2D^2 f(\Omega_m)^2} \quad (7.4)$$

On souhaite à présent transformer les trajectoires de chaque particule  $i$  d'un  $\delta\mathbf{x}_i(D)$ , de telle manière à ce que l'action décroisse. On peut procéder en deux étapes :

1. On minimise d'abord le terme  $\|\boldsymbol{\theta}_i\|^2$  à l'aide d'un transport MAK entre deux instants  $D$  et  $D + dD$ . On affecte alors que le premier terme l'action. Comme celui-ci est toujours positif et que le transport MAK cherche à le minimiser, il va décroître grâce à cette opération.
2. On corrige les trajectoires de manière à ce que l'équation d'évolution

$$\frac{1}{D^{3/2}} \frac{d(D^{3/2}\boldsymbol{\theta}_i)}{dD} = -\kappa(D) \mathbf{grad}_{\mathbf{x}} \phi_g(\mathbf{x}_i, D) \quad (7.5)$$

soit satisfaite. On doit insérer ici la contrainte qu'à l'état initial l'Univers est homogène et que l'on connaît la distribution des particules dans l'état final (mais pas la vitesse). On peut montrer que cette opération ne touche pas au premier ordre de perturbation la valeur de l'action. Pour l'effectuer, on peut discrétiser le temps : on notera  $D_j = j \times \delta D$  la valeur de  $D$  au  $j$ -ème pas de temps de la trajectoire reconstruite. On peut discrétiser (7.5)

$$\begin{aligned} F_{i,j,k} &= \frac{1}{D^{3/2}} \frac{d(D^{3/2}\boldsymbol{\theta}_i)}{dD} \simeq \frac{1}{(\delta D)^2} (x_{i,j+1,k} + x_{i,j-1,k} - 2x_{i,j,k}) + \frac{3}{4(\delta D)D_j} (x_{i,j+1,k} - x_{i,j-1,k}) \\ &= \frac{1}{(\delta D)^2} \left[ \left(1 + \frac{3}{4j}\right) x_{i,j+1,k} + \left(1 - \frac{3}{4j}\right) x_{i,j-1,k} - 2x_{i,j,k} \right] \end{aligned} \quad (7.6)$$

avec  $x_{i,j,k}$  la composante  $k$  de la position de la particule  $i$  à l'instant  $j$  et  $F_{i,j,k} = -\kappa(D) \frac{\partial \phi_g(\mathbf{x}_i, D_j)}{\partial x_k}$ .

On peut mettre cette relation, ainsi que la contrainte à  $D = 0$  et  $D = D_{\text{présent}}$ , sous forme matricielle

$$MX_{i,k} = (\delta D)^2 R_{i,k} \quad (7.7)$$

avec

$$M_{a,b} = \begin{cases} 1 - \frac{3}{4a} & \text{si } b = a - 1 \\ -2 & \text{si } a = b \\ 1 + \frac{3}{4a} & \text{si } b = a + 1 \end{cases}, \quad (7.8)$$

avec  $1 \leq a, b < N$ , et

$$R_{i,k} = \begin{cases} F_{i,j,k} & \text{si } j \neq 0 \text{ et } j \neq N - 1 \\ F_{i,1,k} - \frac{1}{4}x_{i,0,k} & \text{si } j = 1 \\ F_{i,N-1,k} - \frac{5}{4}x_{i,N-1,k} & \text{si } j = N - 1 \end{cases}. \quad (7.9)$$

On impose ici la position des particules aux instants initial et final, tout en appliquant les lois de la dynamique dans l'Univers en expansion. On en déduit alors la position de la particule  $i$  à tout instant :

$$X_{i,k} = (\delta D)^2 M^{-1} R_{i,k} \quad (7.10)$$

Cet algorithme a été implémenté et en ce moment en phase de test. Dans le cas favorable où il améliore les solutions données par MAK, on aurait pour la première fois une méthode capable de reconstruire le régime linéaire de croissance des fluctuations à petite échelle ( $< 12 h^{-1} \text{Mpc}$ ).



## Simulation à petite échelle de la dynamique de la matière-noire

Un deuxième aspect que nous n'avons pas abordé dans cette thèse, mais qui est totalement complémentaire du reste de ce travail, est l'étude de la dynamique de la matière noire à petite échelle. Les simulations à  $N$  corps sont encore incapables de décrire de manière fiable la structure de l'espace des phases. A présent, les simulations à  $N$ -corps fonctionnent suivant le principe qu'il n'est pas nécessaire de suivre précisément la dynamique de chaque particule de matière noire mais qu'il suffit de les étudier par l'intermédiaire d'objet "mesoscopique" pouvant peser au minimum  $\sim 2 \cdot 10^4 M_\odot$  dans le cas de la simulation Via Lactea (Diemand et al., 2007). Certain de ces codes tentent de raffiner cette description. Mais ceux-ci se basent uniquement sur des critères spatiaux (comme la densité en un point à un certain instant), ce qui ne permet pas d'avoir une bonne description de l'espace des phases. Nous proposons ici quelques pistes de recherche pour résoudre ce problème. L'idéal serait de résoudre complètement les équations de Boltzmann-Poisson. Malheureusement, cette méthode est encore largement au-dessus de nos moyens de calculs. Pour des dimensions d'espace des phases inférieures, par exemple pour l'étude de l'effondrement sphérique qui ne nécessite que deux dimensions ( $r$  et le moment cinétique angulaire), la méthode dite du "waterbag" devient faisable : elle consiste à décrire la forme du volume occupé par les particules de matière noire dans l'espace des phases à l'aide de polygones orientés. Néanmoins elle est encore inadaptée, d'un point de vue de la complexité du calcul, à l'étude de la dynamique dans des configurations avec moins de degrés de symétrie. On peut tenter de résoudre ce problème dans le cadre restreint de particules de matières noires *totalemt froides* dans les conditions initiales : nous allons esquisser ce qu'on appelle la méthode de la métrique. Sous l'hypothèse indiquée ci-avant, on peut bien approximer la distribution des particules de matière noire dans l'espace des phases par une nappe tridimensionnelle infiniment fine. La méthode consiste à construire, à l'aide des équations de la dynamique, l'application transformant cette nappe dont la description est triviale dans les conditions initiales à sa description à un instant ultérieur. En résumé, on va chercher à décrire le champ de déplacement en coordonnées lagrangiennes d'une manière consistante : par exemple, en imposant qu'on puisse le décrire localement à l'aide d'un simple développement linéaire. Pour décrire la transformation de cette application avec le temps nous allons nous appuyer sur l'évolution de petits éléments traceurs de ses distortions. Le problème qui se pose est de savoir comment raffiner la description du champ de déplacement dans les cas où la description locale précédente devient insuffisante. Il est possible de le faire en gardant des informations sur l'évolution locale du champ de gravité (comme le champ de marée et ses dérivées spatiales). Grâce à cette information on peut réintégrer les trajectoires de particules tests à partir des conditions initiales jusqu'à l'instant du raffinement. Pour être efficace cette méthode doit :

- raffiner le moins possible tout en gardant l'approximation localement linéaire du champ de déplacement,
- avoir le minimum d'information à garder en mémoire à tout instant,
- avoir le moins de calculs à effectuer lors de la procédure de raffinement.

En fait, ces trois points sont liés. Il est facile de montrer que pour un développement à l'ordre un du champ de gravité on peut satisfaire ces trois contraintes. En effet, à l'ordre un, on cherche à résoudre l'équation suivante :

$$\frac{dx_i}{dt^2} = m_{i,k}(t)x_k \quad (7.11)$$

avec  $x_i$  la  $i$ -ème composante du vecteur  $\mathbf{x}$ ,  $m_{i,k}(t) = \frac{\partial g_i}{\partial x_k}$  représentant le champ de marée du champ de gravité au niveau de l'origine, pour un  $\mathbf{x}$  quelconque. Le théorème des équations différentielles linéaires montre que les  $\mathbf{x}$  solutions se décomposent sur une base de fonctions de dimension trois.<sup>1</sup> Pour avoir l'information complète sur l'application linéaire de transport il suffit donc d'intégrer à chaque pas de temps  $3 \times 6$  éléments qui correspondent à l'évolution des vecteurs de la base des solutions. On peut aussi réduire le plus possible le nombre d'éléments raffinés en introduisant le concept de raffinement anisotrope : on s'attend à ce que le champ de distortion de l'élément de métrique soit assez directif,

<sup>1</sup>On impose que la vitesse est nulle à  $t = 0$  et que l'on cherche les trajectoires dans l'espace des phases.

---

il est donc inutile de raffiner suivant les trois directions de l'espace, une suffit. Pour cela, on peut passer par un système de coordonnées où tous les éléments sont localement isotropes (nous dirons donc “sphériques”). On peut alors en scinder un en deux (créant ainsi deux éléments “fils”) et définir un nouveau système de coordonnées où les deux éléments fils sont aussi localement isotropes. On peut alors transporter les traceurs de la déformation de la métrique après le raffinement grâce à la méthode de transport illustré par l'équation (7.11). Un dernier problème fondamental à résoudre pour que cette méthode marche est d'effectuer le transport jusqu'à l'ordre 2 (ici on ne fait que l'ordre 1), tout en conservant une information compacte telle qu'on vient de le voir pour l'ordre 1. Il est nécessaire d'aller à l'ordre 2 car l'instant où l'on déclenche le raffinement sur un élément est basé sur un défaut à l'ordre 1. Il nous faut donc avoir les bonnes trajectoires des traceurs jusqu'à l'ordre 2 pour éviter d'introduire des erreurs lors de la procédure de raffinement. Ainsi, nous aurions accès pour la première fois à un algorithme de simulation capable de suivre la vraie dynamique de la matière noire jusque dans ses moindres détails à petite échelle. Cela pourrait, par exemple, nous permettre d'identifier des régions dans le ciel où des caustiques, ces flux cohérents et denses de matière noire, se forment. Nous pourrions alors essayer de détecter une quelconque émission dans le domaine des rayons  $\gamma$  afin d'isoler les théories plausibles pouvant expliquer ce qu'est vraiment la matière noire.



# Annexe A Notation utilisée dans le manuscrit de thèse

## A.1 Vecteurs - scalaire

Les vecteurs sont noté en gras. Par exemple le vecteur  $\mathbf{x}$  est noté  $\mathbf{x}$ . Les scalaires sont en italique : le scalaire  $x$ . Le symbole de Kronecker est noté  $\delta_{i,l}$ . Nous rappelons que :

$$\delta_{i,l} = \begin{cases} 1 & \text{si } i = l \\ 0 & \text{sinon} \end{cases} . \quad (\text{A.1})$$

## A.2 Transformée de Fourier

On définit la transformée de Fourier d'une fonction en dimension  $N$  de la manière suivante :

$$\hat{f}(\mathbf{k}) = \int_{\mathbf{x}} \exp(i\mathbf{k} \cdot \mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} = \mathcal{F}(f)(\mathbf{k}) . \quad (\text{A.2})$$

Avec cette définition, la transformée inverse doit alors être

$$f(\mathbf{x}) = \int_{\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{x}) \hat{f}(\mathbf{k}) \frac{d\mathbf{k}}{(2\pi)^N} = \mathcal{F}^{-1}(\hat{f})(\mathbf{x}) , \quad (\text{A.3})$$

afin que  $\mathcal{F}^{-1}(\mathcal{F}(f)) = f$ . Par ailleurs, la convolution  $h$  de deux fonctions scalaires  $f$  et  $g$  s'écrit

$$h(\mathbf{x}) = (f * g)(\mathbf{x}) , \quad (\text{A.4})$$

ce qui donne dans l'espace des modes de Fourier :

$$\hat{h}(\mathbf{k}) = \hat{f}(\mathbf{k}) \hat{g}(\mathbf{k}) . \quad (\text{A.5})$$

## A.3 Projection Aitoff

La projection Aitoff est une projection azimuthale modifiée. On peut effectuer cette projection à l'aide des équations ci-dessous. On pose

$$\alpha = \cos^{-1}(\cos(b) \cos(l/2)) , \quad (\text{A.6})$$

avec  $l$  la longitude (prise entre  $-\pi$  et  $\pi$ ), et  $b$  la latitude. On peut alors construire les coordonnées cartésiennes correspondant à la projection du point sur la sphère ayant les coordonnées  $(l, b)$  :

$$\begin{aligned} x &= \frac{2 \cos b \sin \frac{l}{2}}{\text{sinc } \alpha} \\ y &= \frac{\sin b}{\text{sinc } \alpha} \end{aligned} \quad (\text{A.7})$$

## A.4 Fonction de Bessel sphérique

On utilise la notation  $j_n(x)$  pour les fonctions de Bessel sphériques d'ordre  $n$ . Elles prennent la forme générale suivante :

$$j_n(x) = (-x)^n \left( x \frac{d}{dx} \right)^n \frac{\sin x}{x}. \quad (\text{A.8})$$

On peut écrire explicitement les trois premiers ordres :

$$j_0(x) = \frac{\sin x}{x} \quad (\text{A.9})$$

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} \quad (\text{A.10})$$

$$j_2(x) = \left( \frac{3}{x^2} - 1 \right) \frac{\sin x}{x} - \frac{3 \cos x}{x^2} \quad (\text{A.11})$$

## Annexe B Le problème de minimisation

Nous allons nous intéresser dans ce chapitre à des problèmes purement algorithmiques afin de résoudre l'équation (2.100). Nous limiterons l'exposé au problème plus général de minimization suivant :

$$S_\sigma = \sum_{i=1}^N d(x_i, q_{\sigma i}), \quad \sigma = \arg \min_{\sigma'} S_{\sigma'} \quad , \quad (\text{B.1})$$

avec  $d(a, b)$  un cout d'association de l'élément  $a \in \mathfrak{A}$  à l'élément  $b \in \mathfrak{B}$ .<sup>1</sup> L'algorithme de minimization le plus naïf qu'on puisse inventer est d'évaluer  $S_\sigma$  pour tous les  $\sigma$  possible. Cependant, cela signifie que l'on devra réaliser un nombre de calcul proportionnel au nombre de permutations  $\sigma$ , soit  $N!$ . Si on effectue un nombre fini de calculs par unité de temps, l'algorithme aura une complexité temporelle de  $\mathcal{O}(N!)$ .

### B.1 Un algorithme stochastique

#### B.1.1 Approche dense

Une première amélioration est d'utiliser un algorithme de type stochastique (Croft & Gaztanaga, 1997). On part d'une configuration  $\sigma$  où les particules sont associés aléatoirement, mais toujours bijectivement. Ensuite on effectue en boucle les actions suivantes :

1. On considère la  $n$ -ème itération de l'algorithme.
2. On choisit un élément  $x_i$  au hasard, ainsi qu'un élément  $q_j$ .<sup>2</sup>
3. On trouve les  $j_0$  et  $i_0$  tels que  $j_0 = \sigma_n(i)$  et  $j = \sigma_n(i_0)$ .
4. On crée une nouvelle permutation  $\sigma'_n$  telle que :

$$\sigma'_n(a) = \begin{cases} \sigma_n(a) & \text{si } a \neq i \text{ et } a \neq i_0 \\ j_0 & \text{si } a = i_0 \\ j & \text{si } a = i \end{cases} \quad (\text{B.2})$$

5. Si  $S'_{\sigma} < S_\sigma$ , alors on définit  $\sigma_{n+1} = \sigma'$  pour la prochaine itération  $n+1$ . Sinon on pose  $\sigma_{n+1} = \sigma_n$ .

Cette approche apporte que l'on ne va que se rapprocher du minimum d'une solution. **Il faudra néanmoins que la fonction  $d(a,b)$  soit convexe pour éviter la présence de minima locaux.** **Attention Croft & Gaztanaga (1997) ont montré la convergence pour le cas réduit où  $d(a,b) = \|a - b\|^2$ .** Cependant il est évident que le temps de convergence restera encore très long (quoiqu'inférieur à  $N!$ ).

#### B.1.2 Version sparse

Il y néanmoins une amélioration simple qui permettent de réduire la complexité temporelle de cette algorithme dans le cas où on sait que  $d(x_i, q_j)$  ne peut excéder une certaine valeur  $R_S$ . Elle consiste à changer l'étape 2. de l'algorithme dense de la section B.1.1. On le transforme comme suit : "On choisit un élément  $x_i$  au hasard, ainsi qu'un élément  $q_j$ , tel que  $d(x_i, q_j) < R_S$ ."

---

<sup>1</sup>On supposera dans ce texte que  $|\mathfrak{A}| = |\mathfrak{B}|$ .

<sup>2</sup>Nous verrons un raffinement de cette méthode au prochain paragraphe.

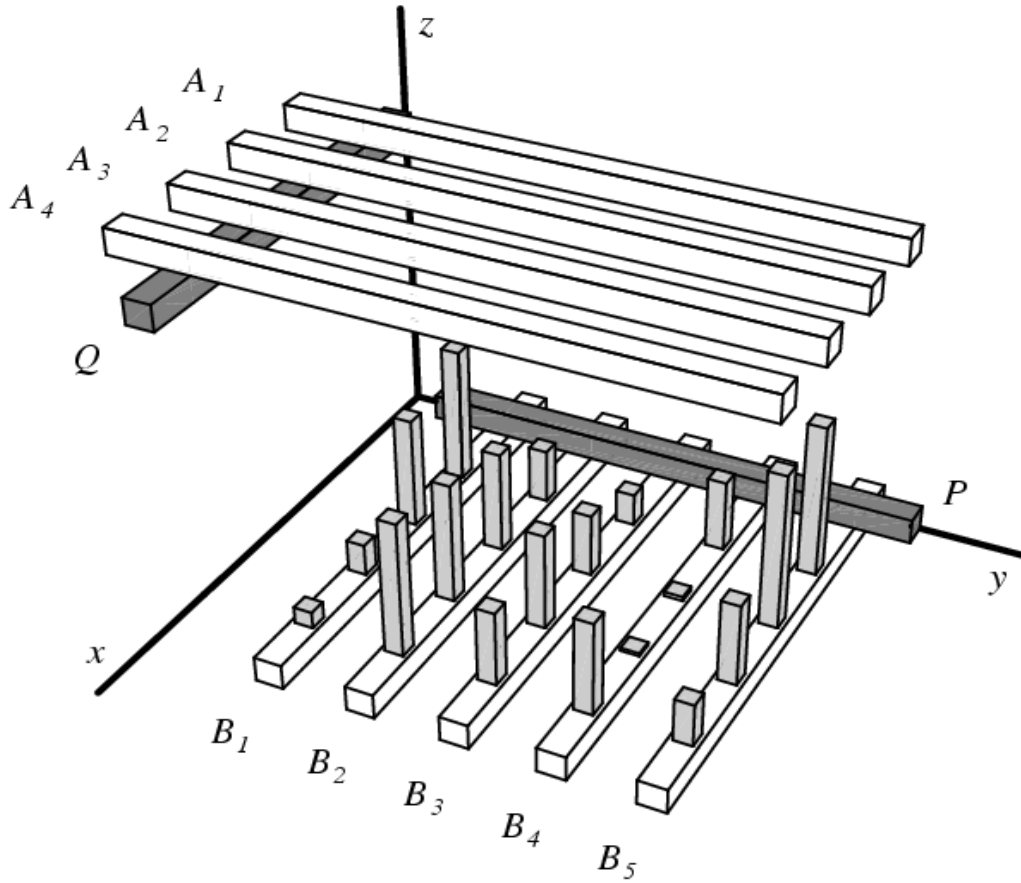


FIG. B.1: *Machine de Hénon* – Les barres horizontales représentent les particules à leur position  $\mathbf{q}$  (en bas) et  $\mathbf{x}$  (en haut), les barres verticales correspondent au coût d’une association. Le système peut évoluer par gravité pour résoudre la problème de la minimisation de  $S_\sigma$ .

## B.2 L’algorithme Auction

### B.2.1 La machine de Hénon et l’idée derrière l’algorithme Auction

L’idée de base de l’algorithme consiste à trouver la permutation  $\sigma$  par itération, les éléments  $x_i$  “luttant” pour acquérir les éléments  $q_j$ . Une idée mécanique d’un tel algorithme existe sous la forme de la machine de Hénon (figure B.1). Hénon (2002) a montré que ce système, sous l’effet des forces de gravité, résout le problème de l’équation (B.1). Cependant il nous faut maintenant une implémentation numérique efficace de cette idée.

### B.2.2 Théorie

Il existe une série d’algorithmes efficaces développés dans un cadre économique par Dimitri Bertsekas (par exemple Bertsekas, 1979, 1998), qu’il a nommé algorithme de Auction (“Vente aux enchères” littéralement). Je vais en rappeler ici les principales étapes.

L’algorithme essaie de faire évoluer l’association  $\sigma$  entre les éléments  $\mathbf{x}_i$  et les  $\mathbf{q}_j$  de telle manière à ce que  $S_\sigma$  soit stationnaire entre deux itérations consécutives de l’algorithme si  $\sigma$  correspond à la solution de (B.1).

Dans Bertsekas (1979), la minimisation du problème (B.1) est obtenue en étudiant le problème dual

de minimisation de pénalités d'association. Il montre en particulier que

$$\min_{\sigma} S_{\sigma} = \min_{p_j: j=1, \dots, n} \left\{ - \sum_j p_j + \sum_i r_i \right\}, \quad (\text{B.3})$$

avec  $a_{i,j} = d(x_i, q_j)$  et

$$r_i = \min_j (a_{i,j} + p_j) \quad (\text{B.4})$$

Une interprétation de l'équation (B.3) est que les  $p_j$  sont des pénalités d'association de l'ensemble des  $x_i$  avec l'élément  $q_j$ . Une fois que l'ensemble  $\{p_j\}$  est déterminé par la maximisation ci-dessus, l'application  $\sigma$  est simplement donné par :

$$\sigma(i) = \arg \min_j \{a_{i,j} + p_j\}. \quad (\text{B.5})$$

Nous allons supposer qu'à l'itération  $n - 1$ , l'algorithme avait déterminé les quantités suivantes :

- l'application  $\sigma' : \mathfrak{A}' \rightarrow \mathfrak{B}'$ , avec  $\mathfrak{A}' \subset \mathfrak{A}$  et  $\mathfrak{B}' \subset \mathfrak{B}$ , avec  $\sigma'$  déterminé par l'équation (B.5). Il s'agit en fait de l'application d'assignation partielle. Cette application ne résout pas le problème minimum global mais propose une bonne approximation. De fait, elle n'est pas forcément définie sur l'ensemble des  $x_i$  mais sur le sous-ensemble  $\mathfrak{A}'$  des  $x_i$ .
- l'ensemble des pénalités sur les éléments de  $\mathfrak{B}'$ , soit  $\{p_j\}$ . Ces pénalités sont calculées de manière à ce que l'application  $\sigma'$  soit bien défini sur  $\mathfrak{A}'$  grâce à la relation (B.5).
- l'ensemble des éléments non-assignés  $\mathfrak{N} = \mathfrak{A}'$ . On souhaite que cet ensemble décroisse régulièrement à chaque itération de l'algorithme.

Chaque itération est composée de deux parties :

- **Phase d'offre** : pour tout élément  $i$  de  $\mathcal{N}$ , les  $x_i$  qui n'ont pas encore trouvé de  $q_j$ , on calcule la quantité  $r_i$  (B.4) et  $A(i)$  comme dans (B.5) avec l'ensemble des pénalités  $\{p_j\}$  calculé à l'itération précédente, ainsi que

$$w_i = \min_{j \neq A(i)} (a_{i,j} + p_j), \quad (\text{B.6})$$

qui est le coût d'association du deuxième meilleur candidat dans l'ensemble complet des  $q_j$ ,  $\mathfrak{B}$ .

- **Phase d'assignation** : Pour chacun des éléments  $i$  de  $\mathfrak{N}$ , l'ensemble des éléments non associés, on effectue les actions suivantes :
  - si l'élément  $q_{j=A(i)}$  candidat est déjà associé à un élément  $x_i$ , c'est-à-dire  $A(i) \in \mathfrak{B}'$ , on pose  $i_0 = \sigma'^{-1}(A(i))$  l'index de l'élément  $x_{i_0}$  auquel est associé  $A(i)$ .
    - Si le “coût effectif” d'association de  $i$  avec  $A(i)$  est inférieur à celui de  $i_0$  avec  $A(i)$ , c'est-à-dire  $r_i < r_{i_0}$ , alors on associe  $i$  à  $A(i)$ ,  $\sigma'(i) = A(i)$  ce qui force  $i_0$  à être retiré de  $\mathfrak{A}'$  et à être mis dans l'ensemble des éléments non-associé  $\mathfrak{N}$  de la prochaine itération. On termine l'assignation en mettant  $i$  dans  $\mathfrak{A}'$  et en le retirant de  $\mathfrak{N}$ .
    - Sinon on n'affecte  $i$  avec personne et on le garde dans  $\mathfrak{N}$ , l'ensemble des éléments  $x_i$  non associés.
  - si l'élément  $q_j$  n'a jamais été associé à un élément  $x_i$ , c'est-à-dire  $A(i) \notin \mathfrak{B}'$ , on ajoute  $i$  à  $\mathfrak{A}'$ , l'ensemble des  $x_i$  ayant été lié avec un  $q_j$ , et l'élément  $q_j$  avec  $j = A(i)$  à  $\mathfrak{B}'$ , l'ensemble des  $q_j$  assigné à un  $x_i$ , et on pose  $\sigma'(i) = A(i)$  pour la prochaine itération.
- **Mise à jour des pénalités** : Les pénalités  $p_j$  sont mises à jour de la manière suivante :

$$p_j \rightarrow \tilde{p}_j = p_j - r_{\sigma'^{-1}(j)} + w_{\sigma'^{-1}(j)} + \epsilon, \quad (\text{B.7})$$

avec  $\epsilon > 0$ . Au passage, on constate que  $p_j$  ne peut qu'augmenter car  $w_i \geq r_i$ .

A chaque itération on est assuré de toujours faire diminuer le nombre d'éléments dans  $\mathfrak{N}$ . Quand son cardinal est égal à 0 on a trouvé une solution du problème à un certain degré d'approximation qui



dépend de la valeur de  $\epsilon$ . Cette solution est, en fait, la même que pour  $\epsilon = 0$  pourvu que  $\epsilon < \epsilon_0/N$  avec

$$\epsilon_0 = \min_{\substack{\{i,j\} \\ a_{i,j} \neq 0}} a_{i,j}. \quad (\text{B.8})$$

Il y a néanmoins un avantage : la présence du  $\epsilon$  évite que la pénalité ne soit pas augmentée lors d'une itération et qu'ainsi on empêche la convergence de l'algorithme.

La complexité temporelle dépend beaucoup sur la façon de faire décroître  $\epsilon$  de sa valeur initiale  $\epsilon_0/N$ . Des expériences numériques montrent qu'une meilleure vitesse de convergence est obtenue si on fait décroître  $\epsilon$  logarithmiquement en 5 itérations.

### B.2.3 Amélioration

Une amélioration substantielle du temps de calcul peut-être obtenue en modifiant la **phase d'offre** en calculant deux quantités supplémentaires pour tout  $i \in \mathcal{A}'$

$$\tau(i) = \arg \min_{j, j \neq A(i)} (a_{i,j} + p_j), \quad (\text{B.9})$$

$$s_i = \min_{j, j \neq A(i), j \neq \tau(i)} (a_{i,j} + p_j). \quad (\text{B.10})$$

A une itération suivante, on peut seulement calculer les coûts effectifs :

$$\alpha_i = a_{i,\sigma(i)} + \tilde{p}_{\sigma(i)}, \quad (\text{B.11})$$

$$\beta_i = a_{i,\tau(i)} + \tilde{p}_{\tau(i)}. \quad (\text{B.12})$$

$$(\text{B.13})$$

Si  $\alpha_i$  et  $\beta_i$  sont toujours inférieurs à  $s_i$  alors on peut éviter de faire la minimisation complète. Sinon, on devra de nouveau procéder à une minimisation complète. En effet, on sait que les pénalités  $p_j \leq \tilde{p}_j$  et donc qu'elles ne peuvent qu'augmenter d'une itération à l'autre. Donc

$$w_i \leq a_{i,j} + p_j \leq a_{i,j} + \tilde{p}_j. \quad (\text{B.14})$$

Si  $\alpha_i \leq w_i$  et  $\beta_i \leq w_i$  alors on sait déjà que, si le minimum a changé, il n'a pu que permuter  $\sigma(i)$  et  $\tau(i)$ . Il ne nous reste donc plus qu'à comparer la valeur de  $\alpha_i$  avec celle de  $\beta_i$ . Dans l'algorithme de la sous-section précédente, on peut modifier la phase d'offre :

- si  $\alpha_i \leq \beta_i$  alors le minimum et le deuxième ne bouge pas. On impose  $r_i = \alpha_i$  et  $w_i = \beta_i$ .
- si  $\alpha_i > \beta_i$ , alors on doit assigner le minimum a changé. On pose  $r_i = \beta_i$ ,  $w_i = \alpha_i$ ,  $A(i) = \tau(i)$  et  $\tau(i)$  est transformé suivant  $\tau(i) \rightarrow \tau(i) = \sigma(i)$ .

L'avantage de procéder ainsi est que, de toute façon, nous devons trouver le minimum et le deuxième minimum. Le surcoût ajouté à noter la position du troisième minimum est faible par rapport au gain de temps de recherche que l'on peut escompter par la modification ci-dessus.

### B.2.4 Implémentation parallèle

Deux types d'implémentation parallèle ont été réalisées :

- une version sur ordinateur à mémoire partagée (architecture type SMP),
- une version sur ordinateur à mémoire non partagée (architecture de type cluster).

L'équilibrage de la charge entre les noeuds se fait lors de la phase d'offre (la plus coûteuse en temps de calcul au début). La construction des  $r_i$ ,  $w_i$  et  $A(i)$  est équirépartie par rapport au nombre d'éléments dans  $\mathfrak{N}$  sur l'ensemble des processeurs de l'ordinateur. Cette approche a l'avantage d'être simple.

Malheureusement il n'existe pas encore de version totalement efficace, c'est-à-dire que tous les processeurs travaillent à tout instant, de cet algorithme sur une architecture parallèle. Ce manque d'efficacité se fait notamment sentir dans les dernières phases de l'algorithme quand il ne reste plus que quelques (voire une seule) particules à assigner. Bertsekas & Tsitsiklis (1997) ne propose pas de solution efficace pour résoudre ce problème.

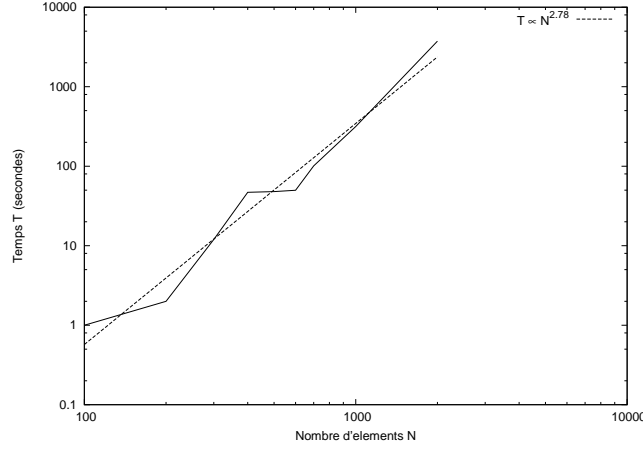


FIG. B.2: *Complexité temporelle de l'algorithme de "Auction"* – On a représenté en ligne solide le temps de calcul nécessaire pour associer  $N$  éléments par paires avec l'algorithme de "Auction". La ligne en tiret représente le meilleur ajustement à cette courbe.

### B.2.5 Version "sparse" – "dense"

Comme dans l'algorithme de Métropolis, la phase de recherche dans l'équation (B.5) et (B.4) peut être accélérée en partant du constat que la distance  $d(x_i, q_\sigma(i))$  est en fait bornée par  $R_S$ . On pourra donc lors de ces minimisations, rechercher uniquement le meilleur élément  $q_j$  auquel  $x_i$  doit être assigné dans une boule de rayon  $R_S$ . Cette modification rend en fait la complexité temporelle de l'algorithme dépendant de la densité  $n$  de  $q_j$ . Dans la pratique, pour les reconstructions nécessitant des corrections de distortion redshift principalement, nous bornons non pas par une sphère mais par un cube dans laquelle la sphère ci-dessus s'inscrit. Pour être en mesure de manipuler toute sorte de forme de grille, nous passons par l'intermédiaire d'une grille grossière indexant les particules  $q_j$  dont la longueur de maille est de l'ordre de  $1/5$  du rayon de recherche. Pour une particule  $x_i$  nous trouvons les voisins  $q_j$  à une distance inférieure à  $R_S$  en considérant les noeuds de la grille grossière se trouvant à moins de 5 unités de grille grossière de la particule  $x_i$ . De plus nous ordonnons les accès aux particules  $q_j$  de la grille grossière de manière à ce que les accès en mémoire centrale soient séquentiels et, ainsi, plus rapides. L'intérêt de la grille grossière est de s'adapter à toutes les situations tout en restant proche de l'algorithme sparse optimal. Comme il est de toute façon nécessaire de considérer un rayon  $R_S$  plus gros que ce qui est strictement nécessaire afin de s'assurer que l'on calcule bien la même chose que l'algorithme dense, on ne devrait pas avoir de perte de performance significative.

### B.2.6 Test de performance

Nous avons testé les performances de l'algorithme en essayant d'assigner des objets entre eux dont les coûts d'association sont générés aléatoirement avec une distribution uniforme dans l'intervalle  $[0 : 10]$ . L'algorithme fonctionne évidemment dans ce cadre là en mode dense et on teste l'efficacité de la méthode "Auction" elle-même. On a représenté sur la figure B.2 un graphe illustrant le temps de calcul nécessaire pour trouver la solution du coût minimal d'association de ces objets par paires. On trouve par une méthode d'ajustement que ce temps croît comme  $O(N^{2.78})$  avec  $N$  le nombre d'objets à associer.



# Annexe C Fonction de corrélations

## C.1 Champ de densité

Le paradigme actuel de la cosmologie stipule que les fluctuations de densité primordiales (et donc à un moindre degré les fluctuations actuelles) correspondent statistiquement à un champ aléatoire gaussien. Ce paradigme est supporté par les observations du fond diffus cosmologique qui montrent qu'essentiellement on a bien affaire à un champ gaussien. Dans ce cas, pour définir entièrement la statistique de ce champ il suffit de connaître la fonction de corrélation à deux points :

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle. \quad (\text{C.1})$$

Cette fonction de corrélation ne dépend en fait que de  $\|\mathbf{r}\|$  car l'Univers est supposé isotrope. On peut aussi exprimer l'équivalent de la fonction de corrélation mais dans l'espace de Fourier qui est la fonction donnant le spectre de puissance des fluctuations :

$$\xi(\mathbf{r}) = \int_{\mathbf{x}} \int_{\mathbf{k}} \int_{\mathbf{k}'} \frac{d^3\mathbf{k} d^3\mathbf{k}' d^3\mathbf{x}}{(2\pi)^6} \exp(-i\mathbf{k} \cdot \mathbf{x}) \exp(-i\mathbf{k}' \cdot (\mathbf{x} + \mathbf{r})) \langle \hat{\delta}(\mathbf{k})\hat{\delta}(\mathbf{k}') \rangle \quad (\text{C.2})$$

$$= \int_{\mathbf{k}} \frac{d^3\mathbf{k}}{(2\pi)^3} \exp(i\mathbf{k} \cdot \mathbf{r}) \langle \hat{\delta}(\mathbf{k})\hat{\delta}(-\mathbf{k}) \rangle \quad (\text{C.3})$$

$$= \int_{\mathbf{k}} \frac{d^3\mathbf{k}}{(2\pi)^3} \exp(i\mathbf{k} \cdot \mathbf{r}) P(k) \quad (\text{C.4})$$

$$= \frac{1}{2\pi^2} \int_{k=0}^{+\infty} dk k^2 P(k) \text{sinc}(k\|\mathbf{r}\|) \quad (\text{C.5})$$

avec

$$P(k) = \langle |\hat{\delta}(\mathbf{k})|^2 \rangle \quad (\text{C.6})$$

et

$$\text{sinc}(x) = \frac{\sin x}{x}. \quad (\text{C.7})$$

$P(k)$  ne dépend que de la norme de  $\mathbf{k}$  pour des raisons d'isotropie de la fonction  $\xi(\mathbf{r})$ . La donnée de  $P(k)$  déterminerait donc toute la statistique du champ de densité. La donnée du spectre de puissance actuel ne correspond pas au spectre de puissance primordial (Tegmark et al., 2004b, 2006). En effet, à cause de la dynamique gravitationnelle non-linéaire le spectre se transforme d'une manière non-triviale. Cependant les paramètres fondamentaux de l'Univers se trouvent dans le spectre de puissance primordial et non dans celui qui contient des non-linéarités. Pour simplifier les conventions, les cosmologistes utilisent la théorie linéaire et normalisent le spectre de puissance des fluctuations à l'instant présent *même si ce spectre ne correspond pas à celui au spectre des fluctuations à l'instant présent*. De plus, le spectre de puissance primordial est en général séparé en deux composantes : la partie "invariance" d'échelle, qui est liée aux "vraies" conditions initiales des fluctuations dans l'univers (et supposée engendrée par une théorie inflationniste), et la partie fonction de transfert  $T(k)$  qui explicite comment la physique de la matière transforme ce spectre dans les premiers instants de l'univers. On écrit donc :

$$P(k) = \mathcal{N} k^n T^2(k) \quad (\text{C.8})$$

avec  $n \simeq 1$ , le spectre des fluctuations primordiales,  $T(k)$  la fonction de transfert de la matière, et  $\mathcal{N}$  la normalisation de ce spectre à redshift nul. Par une méthode de meilleur ajustement on peut obtenir

Type	Meilleur ajustement analytique
Efstathiou et al. (1992)	$T(k) = \left( 1 + \left[ 6.4 \left( \frac{k}{\Gamma} \right) + 3 \left( \frac{k}{\Gamma} \right)^{1.5} + \left( \frac{1.7k}{\Gamma} \right)^2 \right]^{1.13} \right)^{-1/1.13}$ $\Gamma = \Omega_m h$
Bardeen et al. (1986)	$q = \frac{k \left( \frac{T_{\text{CMB}}}{2.7 \text{ K}} \right)^2}{\Omega_m h^2 e \left( -\Omega_B - \sqrt{\frac{h}{0.5} \frac{\Omega_B}{\Omega_m}} \right)}$ $T(k) = \frac{\log_{10}(1 + 2.34q)}{1 + 2.34q} (1 + 3.89q + (16.1q)^2 + (5.46q)^3 + (6.71q)^4)^{-0.25}$
Klypin (2000)	$a_1 = (46.9 \Omega_m h^2)^{0.670} (1 + (32.1 \Omega_m h^2)^{-0.532})$ $a_2 = (12 \Omega_m h^2)^{0.424} (1 + (45 \Omega_m h^2)^{-0.582})$ $\alpha = a_1^{-\Omega_B/\Omega_m} a_2^{-(\Omega_B/\Omega_m)^{0.3}}$ $q = \frac{k \left( \frac{T_{\text{CMB}}}{2.7 \text{ K}} \right)^2}{\Omega_m h^2 \sqrt{\alpha} (1 - \Omega_B/\Omega_m)^{0.6}}$ $T(k) = \frac{\log_{10}(1 + 2.34q)}{1 + 2.34q} (1 + 13q + (10.5q)^2 + (10.4q)^3 + (6.51q)^4)^{-0.25}$

TAB. C.1: Cette table donne quelques meilleur ajustement analytique du spectre de puissance des fluctuations de la matière noire en fonction de paramètres cosmologiques.

différentes fonctions. Nous en présentons deux dans la table C.1. La norme, ici  $\mathcal{N}$ , du spectre de puissance est donnée par les observations sous la forme du paramètre  $\sigma_8$  qui est le taux de fluctuations du champ dans une sphère de  $8 h^{-1} \text{Mpc}$  à l'époque présente. On peut obtenir des statistiques sur les champs de densité lissé en convoluant  $\delta$  avec un noyau  $W$ . On choisira toujours un noyau isotrope (donc qui ne dépend que de la norme de  $\mathbf{x}$ ). La fonction de corrélation des fluctuations lissées est alors :

$$\xi_W(\mathbf{r}) = \langle (\delta * W)(\mathbf{x})(\delta * W)(\mathbf{x} + \mathbf{r}) \rangle \quad (\text{C.9})$$

$$= \int_{k=0}^{+\infty} \frac{dk}{2\pi^2} k^2 \text{sinc}(k||r||) P(k) \hat{W}^2(k) \quad (\text{C.10})$$

Quelques fonctions de filtrages usuelles sont données dans la table C.2. En ce qui concerne la normalisation  $\mathcal{N}$ , il faut utiliser la fenêtre de lissage tophat avec un rayon  $R = 8 h^{-1} \text{Mpc}$ . La variance du champ de densité lissé est alors :

$$\sigma_8^2 = \xi_{\text{TH},8}(0) = \int_{k=0}^{+\infty} dk k^2 P(k) \hat{W}_{\text{TH},8}^2(k), \quad (\text{C.11})$$

où “TH,8” désigne que la quantité sous-jacente a été lissée avec un noyau tophat à  $8 h^{-1} \text{Mpc}$ . Si on connaît  $\sigma_8$  on peut donc en déduire la normalisation :

$$\mathcal{N}(a) = \frac{2\pi^2 \sigma_8^2}{\int_{k=0}^{+\infty} dk k^{n+2} T^2(k) \hat{W}_{\text{TH},8}^2(k)} \quad (\text{C.12})$$

Nom	Expression	Transformée de Fourier	$R_{\text{gauss}}$
Top-hat	$W_R(  \mathbf{x}  ) = \begin{cases} \frac{3}{4\pi R^3} & \text{si }   \mathbf{x}   < R \\ 0 & \text{sinon} \end{cases}$	$\hat{W}_R(  \mathbf{k}  ) = \frac{3}{(kR)^3} (\sin(kR) - kR \cos(kR))$	$\frac{R}{\sqrt{5}}$
Gaussien	$W_R(  x  ) = \frac{1}{(2\pi)^{3/2} R^3} e^{-  \mathbf{x}  ^2/(2R^2)}$	$\hat{W}_R(\mathbf{k}) = e^{-R^2 k^2/2}$	R

TAB. C.2: *Quelques fonctions de filtrage usuelles* – On donne dans cette table le nom de la fonction de filtrage, son expression en espace réel, son expression en espace de fourier, et l'échelle de lissage du filtre gaussien équivalent.

Pour connaître la fonction de corrélation  $\xi(\mathbf{r})$  à un instant quelconque on peut alors utiliser la théorie linéaire, ce qui correspond à simplement multiplier  $\xi$  par un facteur  $D^2(t)/D^2(t_0)$ , avec  $t_0$  désignant l'instant présent et  $D$  le facteur de croissance vu à la section 2.3.2 (page 15).

## C.2 Fonction de corrélation du champ de vitesse lissé

D'après la section 2.4.1 (page 15),  $\theta(\mathbf{x})$  peut s'écrire

$$\theta(\mathbf{x}) = \int_{\mathbf{k}} d^3 \frac{\mathbf{k}}{(2\pi)^3} \frac{i\mathbf{k}}{k^2} e^{i\mathbf{k}\cdot\mathbf{x}} \hat{\delta}(\mathbf{k}) \quad (\text{C.13})$$

Mais encore, si on lisse le champ de vitesse à l'aide du noyau  $W$  :

$$\theta_W(\mathbf{x}) = (\theta * W)(\mathbf{x}) = \int_{\mathbf{k}} \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{i\mathbf{k}}{k^2} e^{i\mathbf{k}\cdot\mathbf{x}} \hat{\delta}(\mathbf{k}) \hat{W}(\mathbf{k}) \quad (\text{C.14})$$

On peut à présent calculer l'autocorrélation de la composante  $\alpha$  champ de vitesse en fonction du spectre de puissance  $P(k)$  introduit dans la section précédente

$$\langle \theta_W^\alpha(\mathbf{x}) \theta_W^\alpha(0) \rangle = \frac{1}{(2\pi)^6} \int_{\mathbf{k}} d^3 \mathbf{k} d^3 \mathbf{k}' \frac{ik^\alpha i{k'}^\alpha}{k^2 {k'}^2} e^{i\mathbf{k}\cdot\mathbf{x}} \langle \hat{\delta}(\mathbf{k}) \hat{\delta}(\mathbf{k}') \rangle \hat{W}^2(\mathbf{k}) \quad (\text{C.15})$$

$$= \frac{1}{3} \frac{1}{(2\pi)^3} \int d^3 \mathbf{k} \frac{1}{k^2} e^{i\mathbf{k}\cdot\mathbf{x}} P(k) \hat{W}^2(\mathbf{k}) \quad (\text{isotropie de la vitesse}) \quad (\text{C.16})$$

$$= \frac{1}{3} \frac{2}{(2\pi)^2} \int_{k=0}^{\infty} dk P(k) \text{sinc}(k||\mathbf{x}||) \hat{W}^2(\mathbf{k}) . \quad (\text{C.17})$$

On peut donc calculer en particulier la variance du champ de vitesse lissé en un point donné

$$\langle (\theta_W^\alpha(0))^2 \rangle = \frac{1}{3} \frac{1}{2\pi^2} \int_0^\infty dk P(k) \hat{W}^2(k) . \quad (\text{C.18})$$

Quelques fonctions de filtrages usuelles sont données dans la table C.2. La représentation de leur transformée de Fourier est donnée dans la figure C.1. Une application intéressante de cette formule est d'avoir une estimation de l'amplitude du champ de vitesse moyen sur un certain volume d'Univers. On a représenté figure C.2 l'évolution de  $\sigma = \beta H \sqrt{\langle (\theta_W^\alpha)^2 \rangle}$  en fonction du rayon de filtrage  $R$ . On voit que la dispersion du champ de vitesse pour une échelle de filtrage  $R$  décroît très faiblement avec  $R$  croissant, et qu'on s'attend à ce qu'une volume d'univers de 100 Mpc/h de rayon se déplace à  $\sim 85 \text{ km s}^{-1}$ . C'est une des conséquences du fait que les mouvements des masses sont sensibles à la distribution de masse à très grande échelle.

On peut tirer un autre enseignement de cette figure. En effet, on peut se demander si l'amplitude de notre mouvement ( $\sim 600 \text{ km s}^{-1}$  par rapport au CMB) est probable dans une cosmologie  $\Lambda\text{CDM}$ . La

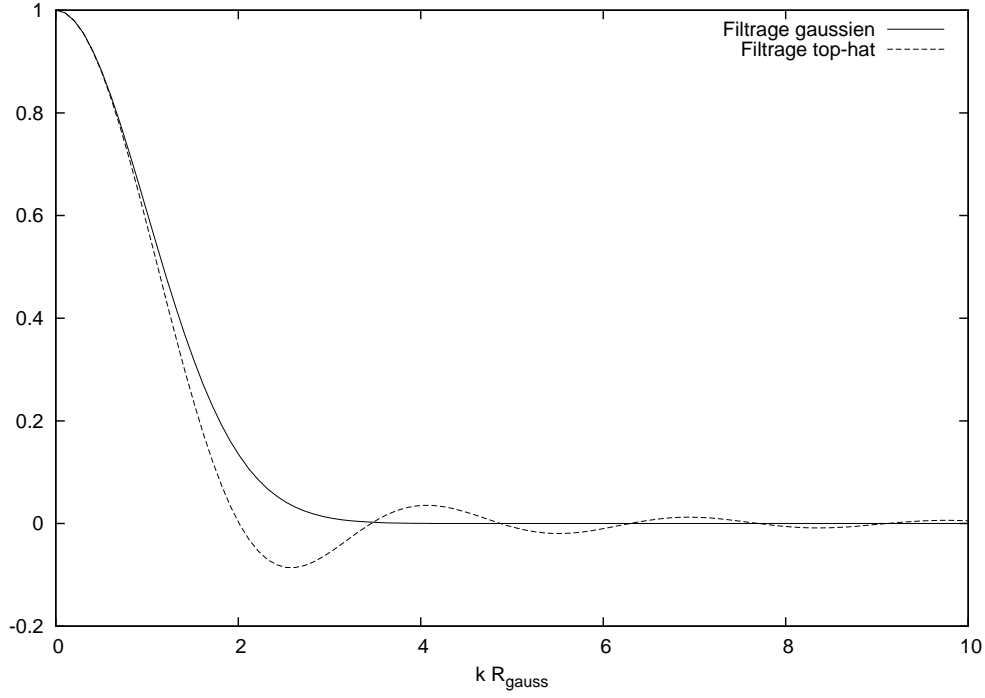


FIG. C.1: Représentation des transformées de Fourier des fonctions de filtrage. Les rayons des différents filtres ont été choisi de manière à ce que les rayons du filtre gaussien équivalent soient identiques ( $R_{\text{gauss}}$ ).

distribution de probabilité de l'amplitude de la vitesse est maxwellienne. Pour une dispersion  $\sigma$  d'une composante du champ de vitesse, on a un maximum de probabilité pour l'amplitude situé à  $V_0 = \sqrt{2}\sigma$ . Autour de cette valeur on peut définir une dispersion :

$$\langle (V - V_0)^2 \rangle = \left( 5 - \frac{8}{\sqrt{\pi}} \right) \sigma^2 \simeq (0.69\sigma)^2. \quad (\text{C.19})$$

Pour un rayon de  $1 \, h^{-1}\text{Mpc}$ , on trouve  $\sigma = 309 \, \text{km s}^{-1}$ . Cela implique une vitesse la plus probable égale à  $V_0 = 436 \, \text{km s}^{-1}$ , mais avec une dispersion de  $213 \, \text{km s}^{-1}$ . Le mouvement du Groupe Local est donc loin d'être improbable, il est tout-à-fait dans la moyenne.

### C.3 Variance de l'estimateur radial du bulk-flow

Nous souhaiterions avoir une idée de l'ordre de grandeur de la différence entre les deux “bulk flow” : le premier provenant de la moyenne du champ de vitesse, l'autre comme l'estimation de cette précédente moyenne à l'aide des vitesses radiales. Dans les graphes de la figure 6.24, nous avons vu que l'essentiel du problème semble venir des effets de projection. En effet, il est difficile d'inférer la valeur de trois composantes d'un champ de vecteur à partir d'une seule. Nous allons essayer de modéliser cela à partir des mesures faites sur le champ de vitesse. Une dérivation plus générale tenant en compte des effets de sélection anisotrope a été obtenue par Regos & Szalay (1989), nous nous contenterons de la version la plus élémentaire pour approcher les concepts. Posons d'abord la quantité :

$$\mathbf{V}_{\text{bulk}} = \frac{1}{V_R} \int_{B_R} d^3\mathbf{r} \, \mathbf{v}(\mathbf{r}) \quad (\text{C.20})$$

le vrai “bulk flow” et

$$U_i = \frac{3}{V_R} \int_{B_R} d^3\mathbf{r} \, \hat{r}_i v_r(\mathbf{r}) \quad (\text{C.21})$$

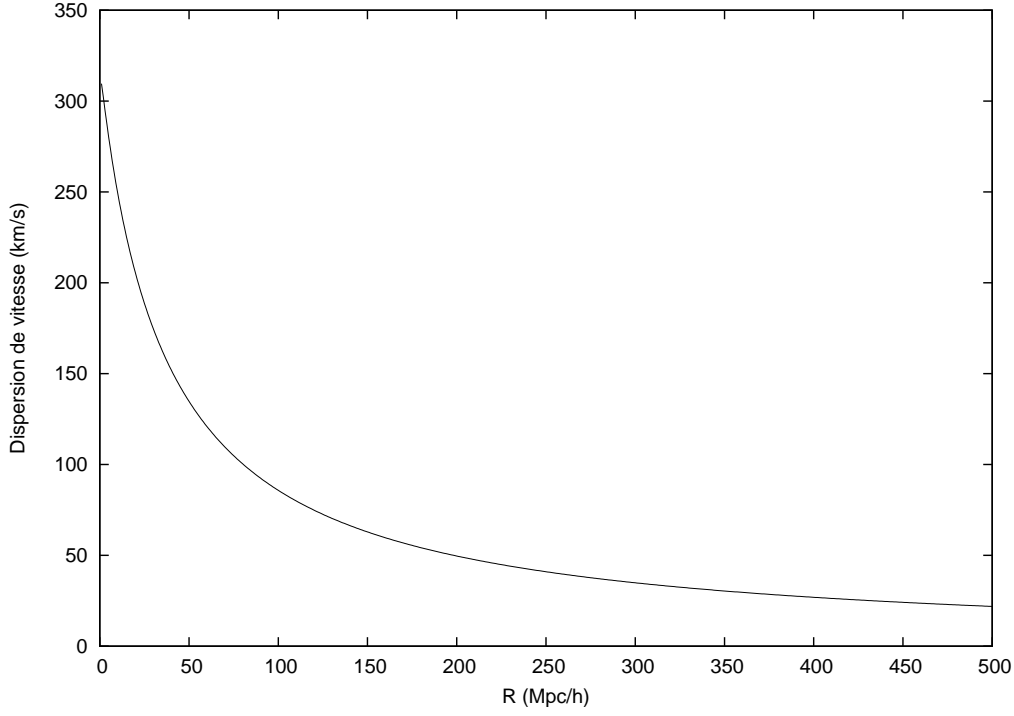


FIG. C.2: *Dispersion du champ de vitesse* – On a représenté ici la dispersion théorique du champ de vitesse lissé avec un fenêtré Top-hat de rayon  $R$ . On a utilisé un spectre de puissance donné par Bardeen et al. (1986), pour une cosmologie  $\Omega_m = 0.30$ ,  $\Omega_\Lambda = 0.70$ ,  $\Omega_K = 0$ ,  $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $n_s = 1$ .

l'estimateur à partir du champ de vitesse radial  $v_r$  du même bulk-flow. Nous souhaitons avoir la valeur de la variance :

$$\sigma_{VU}^2 = \langle (U_i - V_{\text{bulk},i})^2 \rangle = \langle U_i^2 \rangle + \langle V_{\text{bulk},i}^2 \rangle - 2\langle U_i V_{\text{bulk},i} \rangle \quad (\text{C.22})$$

qui est l'écart caractéristique entre  $\mathbf{V}_{\text{bulk}}$  et  $\mathbf{U}_{\text{bulk}}$ . Pour cela reprenons les expressions d'origine :

$$\langle U_i^2 \rangle = \frac{9}{V_R^2} \left\langle \int_{B_R} d^3\mathbf{r} \hat{r}_i \sum_l (\hat{r}_l v_l(\mathbf{r})) \int_{B_R} d^3\mathbf{s} \hat{s}_i \sum_m (\hat{s}_m v_m(\mathbf{s})) \right\rangle \quad (\text{C.23})$$

$$= \frac{9}{V_R^2} \sum_{l,m} \int_{B_R} d^3\mathbf{r} d^3\mathbf{s} \hat{r}_i \hat{s}_i \hat{r}_l \hat{s}_m \langle v_l(\mathbf{r}) v_m(\mathbf{s}) \rangle, \quad (\text{C.24})$$

Si on utilise l'expression de la fonction de corrélation du champ de vitesse C.16 (page 165), pour laquelle on a mis  $W = 1$  (donc pas de lissage) :

$$\langle U_i^2 \rangle = \sum_{l,m} \frac{9}{V_R^2} \sum_l \int_{B_R} d^3\mathbf{r} d^3\mathbf{s} \hat{r}_i \hat{s}_i \hat{r}_l \hat{s}_m \left( \frac{1}{(2\pi)^3} \int d^3\mathbf{k} \frac{P(k)}{k^2} \hat{k}_l \hat{k}_m e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{s})} \right) \quad (\text{C.25})$$

$$= \frac{1}{(2\pi)^3 V_R^2} \int d^3\mathbf{k} \frac{P(k)}{k^2} \sum_{l,m} \hat{k}_l \hat{k}_m \left( \frac{3}{V_R} \int_{B_R} d^3\mathbf{r} \hat{r}_i \hat{r}_l e^{i\mathbf{k} \cdot \mathbf{r}} \right) \left( \frac{3}{V_R} \int_{B_R} d^3\mathbf{s} \hat{s}_i \hat{s}_l e^{-i\mathbf{k} \cdot \mathbf{s}} \right) \quad (\text{C.26})$$

$$= \frac{1}{(2\pi)^3} \int d^3\mathbf{k} \frac{P(k)}{k^2} \sum_{l,m} W_{\text{rad},i,l}(\mathbf{k}) W_{\text{rad},i,m}(\mathbf{k}) \hat{k}_l \hat{k}_m, \quad (\text{C.27})$$



avec

$$W_{\text{rad},i,l}(\mathbf{k}) = \frac{3}{V_R} \int_{B_R} d^3\mathbf{r} \hat{r}_i \hat{r}_l e^{i\mathbf{k}\cdot\mathbf{r}} = \delta_{i,l} (A(kR) + B(kR)) - 3\hat{k}_i \hat{k}_l B(kR) \quad (\text{C.28})$$

$$A(u) = \frac{3}{u^3} \int_0^u v^2 j_0(v) dv \quad (\text{C.29})$$

$$B(u) = \frac{3}{u^3} \int_0^u v^2 j_2(v) dv. \quad (\text{C.30})$$

avec  $j_0$  et  $j_2$  les fonctions de Bessel sphérique d'ordre 0 et 2. L'équation (C.28) est obtenu à partir d'un *ansatz* supposant que  $W_{\text{rad},i,l}$  est de la forme :

$$W_{\text{rad},i,l} = \delta_{i,l} C(\mathbf{k}) + \hat{k}_i \hat{k}_l D(\mathbf{k}), \quad (\text{C.31})$$

on trouve que  $C$  et  $D$  vérifie deux équations linéaires et indépendantes :

$$3C(\mathbf{k}) + D(\mathbf{k}) = \sum_i W_{\text{rad},i,i} = \frac{3}{V_R} \int_{B_R} d^3\mathbf{r} e^{i\mathbf{k}\cdot\mathbf{r}} = 3A(kR), \quad (\text{C.32})$$

$$C(\mathbf{k}) + D(\mathbf{k}) = \sum_{i,l} \hat{k}_i \hat{k}_l W_{\text{rad},i,l} = \frac{3}{V_R} \int_{B_R} d^3\mathbf{r} (\hat{k}\cdot\hat{r})^2 e^{i\mathbf{k}\cdot\hat{r}} = A(kR) - 2B(kR). \quad (\text{C.33})$$

L'avantage est que les deux membres de droite sont totalement invariants par rotation. Donc nous savont que  $C$  et  $D$  ne dépendent que de la norme de  $\mathbf{k}$ . Après quelques manipulations algébriques, on peut montrer que  $C$  et  $D$  prennent la forme donnée par les équations (C.28), (C.29) et (C.30).  $W_{\text{rad},i,l}$  est une fenêtre de lissage anisotrope du champ de vitesse et  $P(k)$  le spectre de puissance des fluctuations du champ de densité de la matière. En constatant que l'expression (C.27) doit être isotrope, cette valeur doit être égal à  $1/3$  de la variance de la norme au carré, qui peut se calculer plus aisément :

$$\langle U_i^2 \rangle = \frac{1}{3} \langle U^2 \rangle = \frac{1}{3(2\pi)^3} \int_{\mathbf{k}} d^3\mathbf{k} \frac{P(k)}{k^2} W_{\text{rad}}(k), \quad (\text{C.34})$$

avec

$$W_{\text{rad}}(k) = \frac{3}{(kR)^3} \int_0^{kR} dv v^2 (j_0(v) - 2j_2(v)) \quad (\text{C.35})$$

On peut écrire de même (cf. C.16) :

$$\langle V_{\text{bulk},i}^2 \rangle = \frac{1}{3(2\pi)^3} \int_{\mathbf{k}} d^3\mathbf{k} \frac{P(k)}{k^2} W_{\text{tophat}}^2(k), \quad (\text{C.36})$$

avec  $W_{\text{tophat}}$  tel que décrit dans le tableau C.2 (page 165). Après quelques manipulations algébriques supplémentaires, on peut obtenir l'expression du troisième terme :

$$\langle V_{\text{bulk},i} U_i \rangle = \frac{1}{3(2\pi)^3} \int_{\mathbf{k}} d^3\mathbf{k} \frac{P(k)}{k^2} W_{\text{rad}}(k) W_{\text{tophat}}(k). \quad (\text{C.37})$$

Cela nous permet donc de calculer la quantité  $\sigma_{VU}$ . On a représenté  $\sigma_{VU}(R)$  pour différentes valeurs du rayon  $R$  sur la figure C.3 pour un univers  $\Omega_{\text{m},0} = 0.30$ ,  $\Omega_{\Lambda} = 0.70$ ,  $\sigma_8 = 1.0$ ,  $h = 0.65$  comme celui de la simulation.

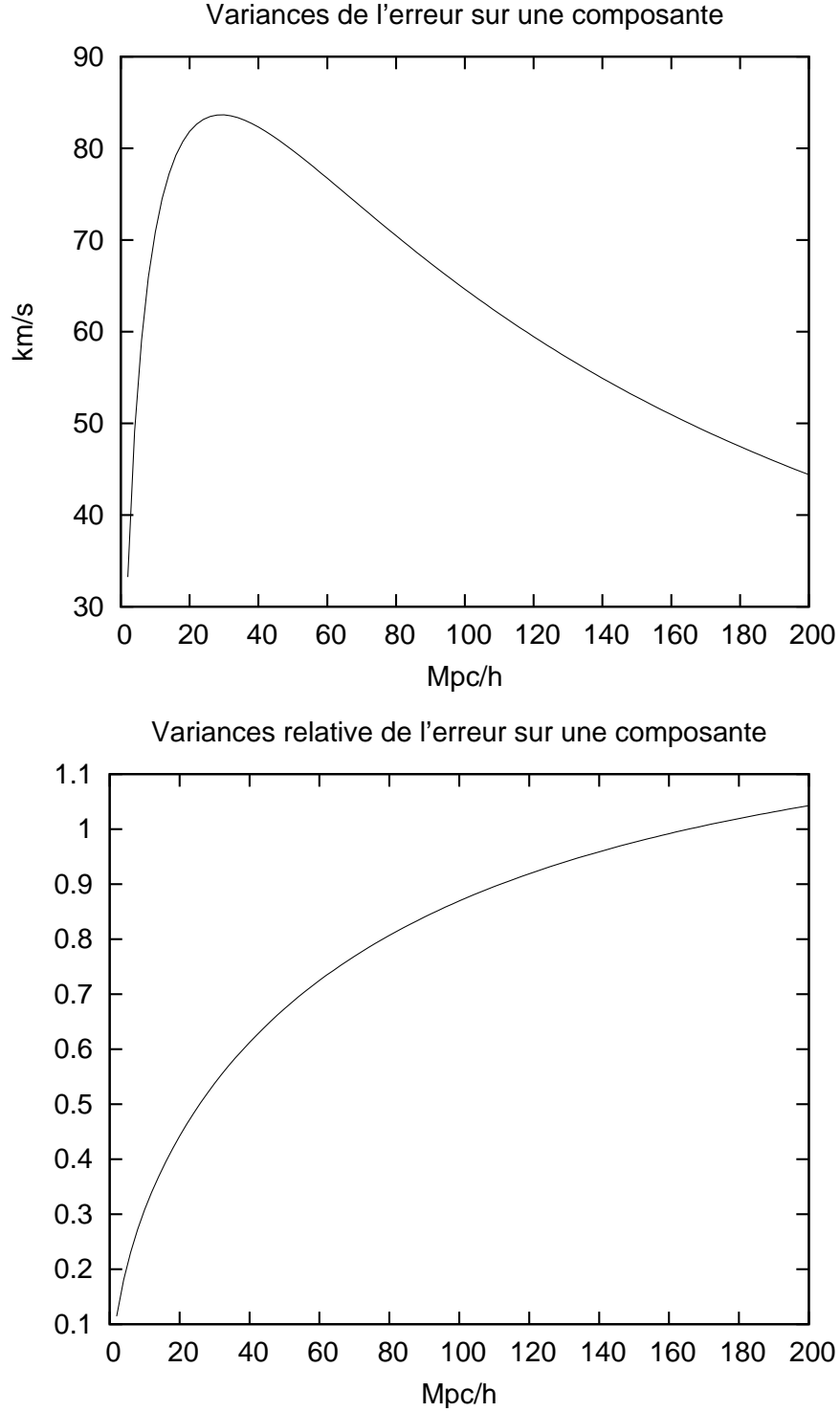


FIG. C.3: *Estimation du bulk-flow, modèle statistique* – On a représenté  $\sigma_{VU}(R)$  pour différentes valeurs du rayon  $R$  sur la figure C.3 pour un univers  $\Omega_{m,0} = 0.30$ ,  $\Omega_{\Lambda} = 0.70$ ,  $\sigma_8 = 1.0$ ,  $h = 0.65$  comme celui de la simulation.



# Annexe D Fonction de masse

## D.1 Présentation de la théorie de Press-Schechter (forme BBKS)

Dans le cadre des champs aléatoire gaussien, Press & Schechter (1974) ont essayé d'établir une prédiction théorique sur la distribution des structures collapsées (type halo de matière noire) en fonction de la masse de ces structures. Cependant la formulation originale (Press & Schechter, 1974) comportait une approximation qui faisait que la théorie était correcte mais à une constante de proportionnalité près. Leur approche a été par la suite corrigée et étendue par Bond et al. (1991). Nous allons maintenant brièvement résumer le formalisme introduit ainsi que les résultats.<sup>1</sup>

Leur théorie est basée sur l'analyse de trajectoire  $\mathcal{T}$  markhovienne dans l'espace (fluctuations, échelle de filtrage). Dans le cadre du modèle de collapse sphérique, on a effondrement total d'une structure (et donc formation d'un halo) quand la densité moyenne de la région considérée passe au-dessus d'un seuil critique  $\delta_{SC}(z)$  qui dépend du redshift  $z$ . Bond et al. (1991) montrent rapidement que l'on peut réexprimer la distribution des objets en fonction de leur masse (soit ici  $Vdn$ ) en terme d'une fonction  $\nu f(\nu)$  qui est reliée à un comptage des trajectoires  $\mathcal{T}$

$$V \frac{dn}{dM} dM = \frac{V\bar{\rho}}{M^2} \frac{d \log \nu}{d \log M} \times \nu f(\nu) \quad (D.1)$$

avec  $\nu = \frac{\delta_{SC}(z)}{\sigma(M)}$ ,  $V \frac{\partial n}{\partial M} dM$  le nombre de halo collapsé dont la masse est entre  $M$  et  $M + dM$  dans un volume  $V$ ,  $\bar{\rho}$  la densité de matière moyenne de l'Univers.  $\sigma(M)$  correspond au taux de fluctuation du champ de fluctuations primordiales si on le lisse sur une échelle donnée par la masse  $M$  (soit  $\propto M/\bar{\rho}$ ). Cette fonction  $f(\nu)$  prend une forme simple dans le cadre du modèle de collapse sphérique analysée par Bond et al. (1991)

$$\nu f(\nu) = 2 \left( \frac{\nu^2}{2\pi} \right)^{1/2} \exp \left( -\frac{\nu^2}{2} \right) \quad (D.2)$$

avec  $\int_0^{+\infty} f(\nu) d\nu = 1$ . Cette formulation bien que donnant des résultats déjà impressionnant au vu des résultats des simulations est néanmoins insuffisante pour décrire l'intégralité de la fonction de masse.

## D.2 La modification de Sheth, Mo & Tormen

Sheth et al. (2001) [et plus tard Sheth & Tormen (2002)] ont effectué une modification du formalisme précédant en arguant du fait que l'on avait négligé les effets du collapse ellipsoïde sur la fonction de masse. Une modélisation statistique du collapse ellipsoïdal donne<sup>2</sup>

$$\nu f(\nu) = 2A (1 + \nu'^{-2q}) \left( \frac{\nu'^2}{2\pi} \right)^{1/2} \exp \left( -\frac{\nu'^2}{2} \right) \quad (D.3)$$

avec  $A$  tel que  $\int_0^{+\infty} f(\nu) d\nu = 1$ ,  $A \simeq 0.322$ ,  $a = 0.707$ ,  $q = 0.3$ .

<sup>1</sup>Nous renvoyons le lecteur aux articles sus-cités pour une description complète du formalisme.

<sup>2</sup>En fait, il s'agit d'une modélisation statistique suivi d'un tirage aléatoire de trajectoire markhovienne dans un champ gaussien.

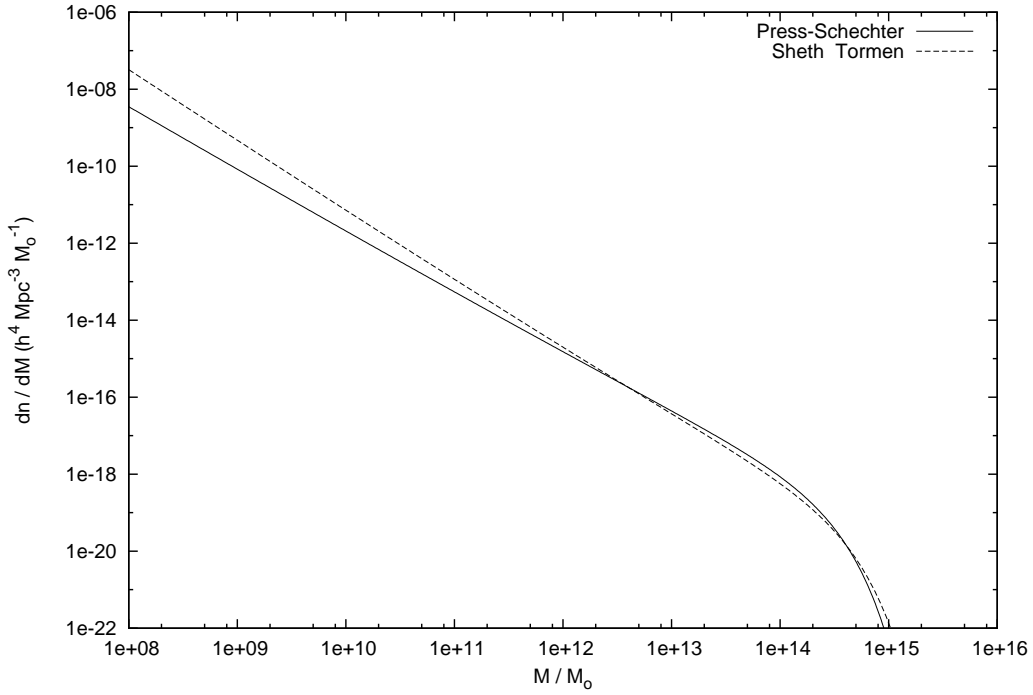


FIG. D.1: *Fonction de masse de Press & Schechter et de Sheth & Tormen* – On a représenté ici pour la même cosmologie et le même spectre de puissance les deux fonctions de masse prédites par, d’une part, la théorie de Press & Schechter, d’autre part, la modification de Sheth & Tormen.

Un meilleur ajustement sur la simulation GIF (Kauffmann et al., 1999) suggère une modification de la paramétrisation suivante :

$$\nu f(\nu) = 2A (1 + \nu'^{-2q}) \left( \frac{\nu'^2}{2\pi} \right)^{1/2} \exp \left( -\frac{\nu'^2}{2} \right) \quad (\text{D.4})$$

avec  $\nu' = \sqrt{a}\nu$  ; On a représenté sur la figure D.1, les deux fonctions de masse.

### D.3 Compter la fraction manquante

On souhaite estimer la fraction de masse présente dans des grumeaux de matière noire de masse supérieur à un certain seuil  $M_0$ . Une telle estimation nous permettrait de savoir de combien la méthode de reconstruction MAK peut être biaisé par la coupure en luminosité du catalogue utilisé pour la reconstruction. On supposera dans le cas de catalogues de galaxies que la luminosité de celles-ci est un bon traceur de la masse du halo de matière noire qui l’abrite. Ainsi, à un rapport  $M/L$  près, on a accès à la coupure en masse du catalogue. Pour évaluer la fraction de masse qu’on observe, on peut utiliser la fonction de masse Sheth & Tormen (2002) (équation (D.4)) :

$$M(> M_0) = \int_{M_0}^{+\infty} dM n(M) M \quad (\text{D.5})$$

On a donc la fraction de masse  $f(M > M_0)$

$$f_{M>M_0} = \frac{M(> M_0)}{M(> 0)} = \int_{M_0}^{+\infty} \frac{dM}{\rho_0} \frac{dn}{dM} M = \int_{\log M_0}^{+\infty} \frac{d(\log M)}{\rho_0} \frac{dn}{dM} M^2. \quad (\text{D.6})$$

La fraction de masse inférieure au seuil est aussi décrite par :

$$f_{M < M_0} = 1 - f_{M > M_0} \quad (\text{D.7})$$

**Considerations numériques** La fonction de masse  $n(M)$  diverge à petite masse comme  $1/\sqrt{M}$ . Cela peut être un problème quand on fait une évaluation numérique de l'intégrale l'équation (D.6). Il est plus aisé de calculer  $f(M > M_0)$  que  $f(M < M_0)$ . Cette fonction a été représentée sur la figure 5.1 (page 70).



## Annexe E Filtrage adaptatif

Dans beaucoup de cas, des champs sont échantillonnés par des particules qui ne sont pas régulièrement disposées sur une grille, qui présente des propriétés intéressantes pour le calcul de quantités dérivées ou intégrées du champ. On peut par exemple penser naturellement au champ de vitesse d’une simulation. Celui-ci est échantillonné par l’intermédiaire des propriétés physiques des particules de matière dans une simulation cosmologique, néanmoins pour lui donner une forme “utilisable” il est nécessaire de le projeter sur une grille. Un autre champ qui est intéressant est la densité de probabilité bidimensionnelle  $P(v_{\text{sim}}, v_{\text{rec}})$  qui est développé dans la section 4.2.1. Dans ce cas, il est nécessaire, à partir d’une distribution de points statistiquement équivalents, de produire un champ lisse décrivant la distribution de probabilité de trouver un point en un certain endroit du diagramme  $(v_{\text{sim}}, v_{\text{rec}})$ . Nous allons voir ici une méthode basée sur les algorithmes de lissage dits “SPH” (Smoothed Particle Hydrodynamics) pour rééchantillonner ces champs d’une manière “optimale”.

### E.1 Le lissage adaptatif général

La méthode SPH propose de calculer la densité d’un champs  $\tilde{A}$  en un point  $\mathbf{r}$  d’un espace à  $d$ -dimension à partir de la valeur du champ  $A_i$  véhiculée par une particule  $i$ . Pour cela on peut écrire de manière tout-à-fait générale :

$$\tilde{A}(\mathbf{r}) = \sum_{i=1}^N A_i W_i(\mathbf{r}), \quad (\text{E.1})$$

qui est l’équivalent d’un filtrage linéaire en  $A_i$ , où  $W_i$  est à définir, et dépend en général d’une manière non-triviale de  $\mathbf{r}$  et  $i$ .

Dans la méthode traditionnelle de lissage SPH, on considère qu’une particule trace un nuage de taille  $2R_{\text{SPH}}$ , où  $R_{\text{SPH}}$  dépend de la densité locale, qui correspond à la distance caractéristique entre cette particule et ses  $N_{\text{SPH}}$  plus proches voisins. On trouve alors les  $W_i$  correspondant. Le problème de cette approche est que l’on peut laisser des régions où le champ estimé n’est pas défini. Une approche alternative (Colombi et al., 2007) est de considérer que les noeuds de la grille elle-même peuvent être considérée comme des particules virtuelles, pour lesquelles nous cherchons les  $N_{\text{SPH}}$  plus proches voisins pour construire le noyau SPH qui lui est associé. Il faut cependant noter que l’intégration se fait dans un petit volume  $\Delta r^d$  et toutes les particules appartenant au volume d’un des sites de grille doivent participer à cette intégrale. Ceci peut être résolu d’une manière approchée de la manière suivante :

1. on compte et on enregistre, pour chaque site de grille  $p$ , le nombre  $N_p$  de particules qu’il contient.
2. si  $N_p > N_{\text{SPH}}$ , alors on effectue l’interpolation SPH sur le site de grille comme il va être expliquée par la suite mais avec  $N_p$  particules contributrices à la place des  $N_{\text{SPH}}$ .
3. si  $N_p \leq N_{\text{SPH}}$ , alors on cherche pour le site de grille les  $N_{\text{SPH}}$  plus proche voisins et on calcule l’interpolation SPH sur le site de grille comme on va l’expliquer.

Nous allons voir à présent deux types de lissage qui peuvent être effectué suivant le la forme précise de  $W_i$  que nous allons prendre.



## E.2 Le lissage avec conservation de l'intégrale du champ

### E.2.1 Formalisme général

On peut penser à un premier lissage où l'on cherche à conserver la somme

$$\sum_{l=1}^N A_l \quad (\text{E.2})$$

Nous allons proposer un lissage SPH qui présente cette propriétés très intéressante. Pour cela nous partons de l'expression du lissage SPH "standard" sur une grille pour lequel on se donne un noyau  $\mathcal{S}(x)$ . Au point de coordonnées  $j_1, \dots, j_d$ , on a la quantité interpolée :

$$\tilde{A}(j_1, \dots, j_d) = \frac{1}{[R_{\text{SPH}}(j_1, \dots, j_d)]^d} \sum_{l=1}^N A_l Q_l \mathcal{S}\left(\frac{d_l(j_1, \dots, j_d)}{R_{\text{SPH}}(j_1, \dots, j_d)}\right), \quad (\text{E.3})$$

avec  $Q_l$  un poids donné à la particule  $l$  pour s'assurer de la conservation de la quantité (E.2),  $d_l(j_1, \dots, j_d)$  la distance de la particule  $l$  au site de grille  $(j_1, \dots, j_d)$ . Si on renforce cette conservation on obtient :

$$\sum_{l=1}^N A_l = \sum_{j_1, \dots, j_d} \tilde{A}(j_1, \dots, j_d) \quad (\text{E.4})$$

$$= \sum_{j_1, \dots, j_d} \frac{1}{[R_{\text{SPH}}(j_1, \dots, j_d)]^d} \sum_{l=1}^N A_l Q_l \mathcal{S}\left(\frac{d_l(j_1, \dots, j_d)}{R_{\text{SPH}}(j_1, \dots, j_d)}\right) \quad (\text{E.5})$$

$$= \sum_{j_1, \dots, j_d} \sum_{l=1}^N A_l Q_l \frac{1}{[R_{\text{SPH}}(j_1, \dots, j_d)]^d} \mathcal{S}\left(\frac{d_l(j_1, \dots, j_d)}{R_{\text{SPH}}(j_1, \dots, j_d)}\right) \quad (\text{E.6})$$

$$= \sum_{l=1}^N A_l Q_l \sum_{j_1, \dots, j_d} \frac{1}{[R_{\text{SPH}}(j_1, \dots, j_d)]^d} \mathcal{S}\left(\frac{d_l(j_1, \dots, j_d)}{R_{\text{SPH}}(j_1, \dots, j_d)}\right). \quad (\text{E.7})$$

Une condition nécessaire et suffisante pour avoir conservation pour toutes les configurations de particules possibles et pour tous les  $A_l$  est de choisir la valeur suivante pour  $Q_l$  :

$$Q_l = \left[ \sum_{j_1, \dots, j_d} \frac{1}{[R_{\text{SPH}}(j_1, \dots, j_d)]^d} \mathcal{S}\left(\frac{d_l(j_1, \dots, j_d)}{R_{\text{SPH}}(j_1, \dots, j_d)}\right) \right]^{-1} \quad (\text{E.8})$$

Nous utiliserons par la suite un choix particulier de noyau  $\mathcal{S}(\mathbf{x})$  (Monaghan, 1992) :

$$\mathcal{S}(x) \propto \begin{cases} 1 - 1.5x + 0.75x^3, & 0 \leq x < 1 \\ 0.25(2 - x)^3, & 1 \leq x < 2 \\ 0 & \text{sinon} \end{cases} \quad (\text{E.9})$$

qui est une B-spline cubique. Ce noyau présente l'avantage d'être à support compact, d'être continu, d'avoir une dérivée continue. Le champ interpolé sera donc au moins continu et à dérivée continue.

### E.2.2 Application à l'interpolation du champ de vitesse et de la masse

**Détails théoriques** Nous nous plaçons à présent dans l'espace réel et par conséquent  $d = 3$ . La principale difficulté est de reconstruire un champ de vitesse lisse sur une grille régulière. Le champ de vitesse que l'on cherche à évaluer est une quantité purement eulérienne. Nous définissons le champ de

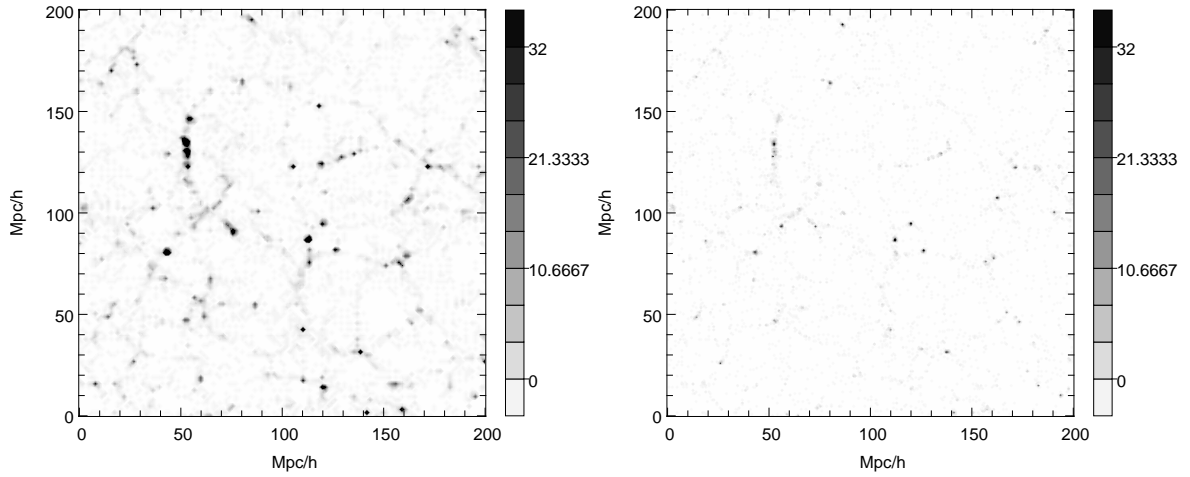


FIG. E.1: *Nombre de particules par sites de grille* – Ces deux graphiques donnent le nombre de particules présentes sur un volume de sites de grille, dans un plan pris au milieu de la simulation. Le graphique de gauche a été produit pour une grille de résolution  $128^3$ , le graphique de droite pour une grille de résolution  $256^3$ .

densité de matière et le champ de vitesse à l'aide de la fonction de distribution de particules identiques dans l'espace des phases  $f(\mathbf{r}, \mathbf{v}_L)$ . Dans ce cas, les quantités eulériennes moyenne s'écrivent :

$$\rho(\mathbf{r}) = \int d^3\mathbf{v}_L f(\mathbf{r}, \mathbf{v}_L), \quad (\text{E.10})$$

pour le champ de densité, et

$$\mathbf{v}(\mathbf{r}) = \frac{1}{\rho(\mathbf{r})} \int d^3\mathbf{v}_L \mathbf{v}_L f(\mathbf{r}, \mathbf{v}_L), \quad (\text{E.11})$$

pour le champ de vitesse.

Nous allons estimer le champ de densité de matière  $\tilde{\rho}(j_1, j_2, j_3)$  s'écrit à l'aide de ce formalisme, avec  $A_l = m_l$  la masse des particules traceuses, comme :

$$\tilde{\rho}(j_1, j_2, j_3) = \frac{\tilde{m}(j_1, j_2, j_3)}{\Delta r^3}. \quad (\text{E.12})$$

De même on peut calculer le champ de vitesse interpolé :

$$\mathbf{v}(j_1, j_2, j_3) = \frac{\tilde{\mathbf{p}}(j_1, j_2, j_3)}{\tilde{m}(j_1, j_2, j_3)} \quad (\text{E.13})$$

avec  $\tilde{\mathbf{p}}(j_1, j_2, j_3)$  le champ d'impulsion interpolé sur la grille à partir des quantités  $p_l = m_l \mathbf{v}_l$  des particules.

**Application numérique** On a représenté sur la figure E.1 le nombre de particules par site de grille pour deux résolutions de grille. On voit que même pour une résolution de  $128^3$  le volume occupée par les sites de grille dont le nombre de particules voisines excède 32 est petit (points noirs). Pour une grille de  $256^3$ , ceux-ci deviennent négligeables. On peut donc en fait travailler avec une version simplifiée de l'algorithme où on ne se préoccupe pas du nombre de particules sur un site de grilles, mais où on prend toujours les 32 particules les plus voisines, pourvu que la grille ait une résolution suffisante comme dans le cas de la grille  $256^3$ . Nous avons représenté le champ de densité calculé de cette manière pour la simulation dans une boîte de  $200 h^{-1} \text{Mpc}$  de côté à l'aide d'une grille  $256^3$  sur la figure E.2.

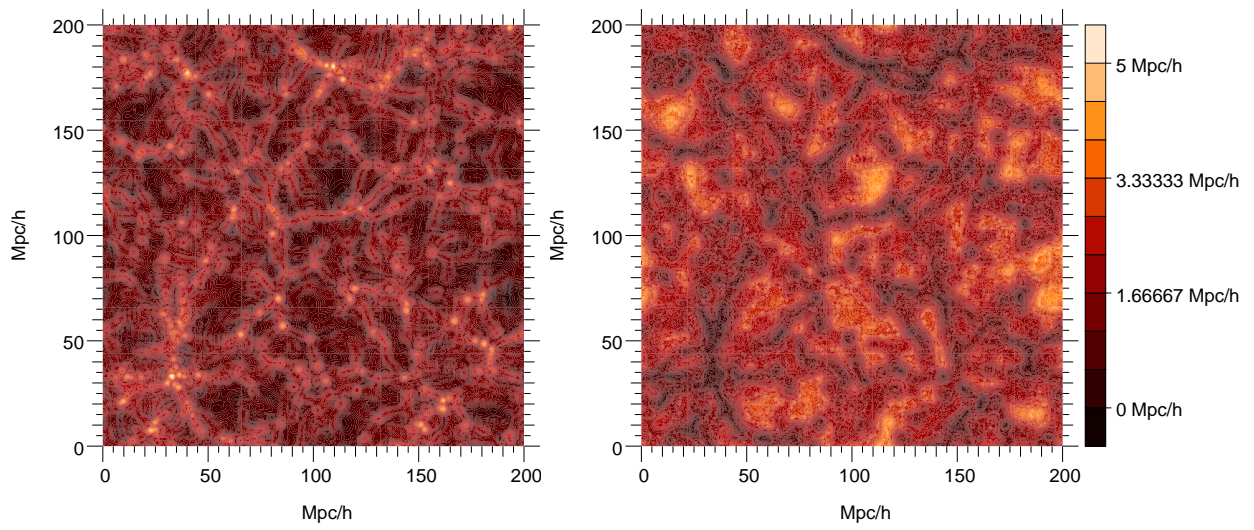


FIG. E.2: *Illustration du lissage adaptatif pour l'obtention d'un champ de densité d'une simulation* – On a représenté sur le panneau de gauche le résultat du filtrage adaptatif du champ de densité (échelle de couleur proportionnelle au logarithme de la densité). Sur le panneau de droite, on a représenté l'échelle de lissage choisit par le filtre pour chaque point de la grille. Le lissage peut aller jusqu'à  $5 h^{-1}\text{Mpc}$  de rayon.

# Annexe F Proceedings

SF2A 2006

D. Barret, F. Casoli, T. Contini, G. Lagache, A. Lecavelier, and L. Pagani (eds)

## RECONSTRUCTION OF THE PRIMORDIAL DENSITY FIELD FROM REDSHIFT AND DISTANCE CATALOGS OF GALAXIES

G. Lavaux<sup>1</sup>, R. Mohayaee, R.B. Tully<sup>2</sup>, S. Colombi and J. Silk<sup>3</sup>

**Abstract.** Peculiar velocity reconstruction methods allow one to have a deeper insight into the distribution of dark matter: both to measure  $\Omega_m$  and to characterize the primordial density fluctuations. We present here the Monge-Ampère-Kantorovitch method applied to redshift reconstruction. We show what are the limitations and the problems to overcome to prevent systematic errors. This leads for the first time to a good representation of the peculiar velocity field within a 8,000 km/s sphere. We measure  $\Omega_m = 0.237^{+0.063}_{-0.056}$  using a catalog of distances within 3,000 km/s, which is in agreement with the analysis of WMAP and SDSS data.

### 1 Introduction

In the last 20 years, large surveys have given us the redshift positions of nearly a million galaxies. However, these differ from the real position by the radial peculiar velocities which are an important probe of the distribution of dark matter. Peculiar velocities are difficult to determine above 2000 km/s because measurement errors increase linearly with redshift. Here, we show how they can be accurately reconstructed from redshift positions and, in doing so, we also put constraints on the dark matter density. Our method also gives the primordial density fluctuations above 5 Mpc/h with no extra approximation. The reconstructed primordial universe corresponds exactly to our own, and not just in a statistical sense as is the case with CMB data.

To achieve our tasks we need find the mapping function between Lagrangian ( $\vec{q}$ ) and Eulerian ( $\vec{x}$ ) coordinates. The orbit of galaxies, taken as mass-tracers, between these coordinates are given by the stationary points of the Euler-Lagrange action. This problem, as formulated by Jim Peebles in 1989, does not have a unique solution, probably because of wrong choice boundary conditions and/or multistreaming. However above a few Mpc multistreaming is marginal and we shall see with the choice of proper boundaries, a unique solution can be achieved.

### 2 Reconstruction methods

We use the Monge - Ampère - Kantorovitch (MAK) method. It is a simplification of the exact problem proposed by Peebles and is based on the hypotheses that the displacement field is derived from a convex potential and constrained by the local mass conservation. These hypotheses are supported by simulations and arise naturally within Lagrangian perturbation theory. It has been proved (Brenier et al. 2003) that solving this problem is equivalent to solving a minimization problem:

$$I_\sigma = \min_{\sigma \in \mathbf{S}^N} \sum_{i=0}^N (\vec{x}_i - \vec{q}_{\sigma(i)})^2 \quad (2.1)$$

where  $i$  runs over particles of the same mass,  $\mathbf{S}^N$  is the group of permutations of rank  $N$ . Solving this problem directly is technically difficult. A fast algorithm, called the auction algorithm, has been invented by Bertsekas (1988), which has an effective complexity of  $N^{2.25}$  where  $N$  is the number of particles.

This method has been tested directly on N-body simulations (Mohayaee et al. 2005). The difference between the reconstructed density field and the original one has a relative dispersion of 28% and a correlation coefficient of 96% when both are smoothed at 5 Mpc/h.

<sup>1</sup> IAP, 98bis boulevard Arago, 75014 Paris, France

<sup>2</sup> Institute for Astronomy, University of Hawaii, Honolulu, HI 96822, U.S.A.

<sup>3</sup> Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom

### 3 Correction of errors through the analysis of mock catalogs

In general, catalogs of observations always have errors. We present here a few important errors and discuss in more detail three of them in the next subsections.

- **Boundary effects:** The *Lagrangian volume* ( $\bar{q}$ ) needed for our reconstruction (see Eqn 2.1) is not given by catalogs. Determination of this unknown quantity can cause errors as discussed below.
- **Incompleteness:** Catalogs are built on survey based on the apparent luminosities of objects: farther intrinsically more luminous objects will look fainter and hence be absent from the catalog (discussed below).
- **Redshift distortion:** In a catalog, distance of an object is generally expressed in redshift. This is not the true distance because it also includes the projection of its peculiar velocity along the line of sight. That is why we tend to use boundary constraints in redshift and not in real space. On the other hand, the reconstruction is expressed in terms of real coordinates and hence one needs make an appropriate correction for the effect.
- **Mass-to-light ratio:** Masses are not measured directly in the catalog. They are estimated either from the local dynamics of compound objects (such as groups) or from a general extrapolation of the mass to luminosity function applied to the luminosity of the object. We can cross-correlate the reconstructed peculiar velocities with the measured distances to determine the best mass-to-light ratio.
- **Zone of avoidance:** Catalogs generally do not have any data in the *galactic plane*. To fix this problem, we assume the general statistical properties are conserved across the boundary of the obscured region and hence fill it accordingly. This enables us to be able to keep the reconstruction of other parts of the catalog intact.
- **Malmquist bias:** To estimate the cosmological parameters, we compare the reconstructed displacement field to the measured distances which are affected by log-normally distributed errors. This effect is known as the volume malmquist bias, which we have taken into account by carefully computing the likelihood function for the parameters.
- **Modeling error:** The method uses an approximation of the dynamic to recover the peculiar velocity field. A measurement on mock catalogs show that the error on peculiar velocities has a Lorentzian distribution and we take this into account in the determination of  $\Omega_m$ .

#### 3.1 Redshift distortion

The easiest way to correct for redshift distortion is to assume the Zel'dovich approximation to compute the current peculiar velocity of catalog objects assuming some displacement vector. If we use this hypothesis we can build a new cost function from Eq. 2.1:

$$I_{\sigma,s} = \sum_{i=0}^N \left( (\vec{s}_i - \vec{q}_{\sigma(i)})^2 - \frac{\beta(2+\beta)}{1+\beta} \frac{((\vec{s}_i - \vec{q}_{\sigma(i)}) \cdot \vec{s}_i)^2}{\|\vec{s}_i\|^2} \right) \quad (3.1)$$

with  $\beta = \frac{\dot{D}}{HD} \approx \Omega_m^{5/9}$ ,  $D$  the growth factor,  $H$  the hubble constant and  $\vec{s}_i$  the redshift position of the  $i$ -th object.

The problem is that we make a linear approximation to recover the peculiar velocity and that we are again using this approximation to get the resulting predicted redshift distortion by MAK. This tends to sum up the errors and decrease the amount of useful signal.

We must be aware of two major problems. First, cluster fingers of god in redshift catalogs must be collapsed to get a good reconstruction. Second, redshift distortion generally introduces a loss of real space connectivity. This effect is particularly important on the boundaries where massive groups or clusters of galaxies may be separated from the main catalog in real space though they are connected in redshift space.

#### 3.2 Lagrangian volume

It is possible to minimize the effect of our ignorance of the initial Lagrangian volume on the central part of the catalog. If we put the inhomogeneous catalog inside a bigger homogeneous cube and we do the reconstruction

using a cubic Lagrangian volume of the same size, then we will exactly resolve the problem concerning the cube. However, as the boundaries for the catalog are not correct, the reconstruction will be fuzzy in the outer parts and at the same time the central part will be unaffected. This has been tested on mock catalogs and on average one recovers the density field with a relative dispersion of 42% and a correlation coefficient of 90%.

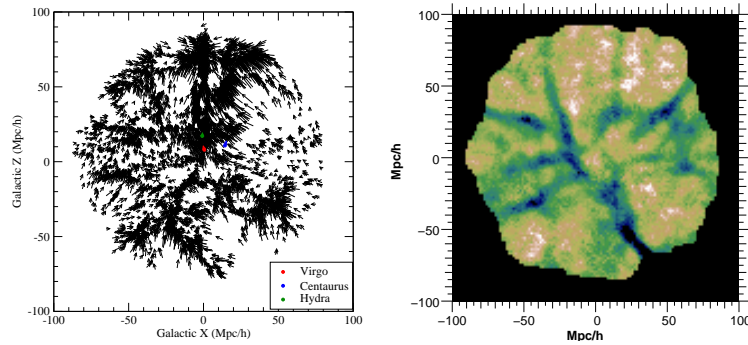
### 3.3 Catalog incompleteness

All catalogs suffer from incompleteness caused mainly by the decrease in luminosity of objects at large distances. To account for the missing luminosity at a given distance, we boost the luminosities of the object at that distance by an amount given by the Schechter luminosity function. This is not a trivial correction and probably will change the dynamics because too much mass can be assigned to the observed objects and thus these objects will be displaced less than the one whose luminosity has not been boosted. The conclusion of the study on mock catalogs is that the higher resolution part of the catalog is generally not affected by random uncertainties in the lower resolution part.

## 4 Application to real data

### 4.1 Velocity field

A direct application of the MAK reconstruction method on a redshift catalog gives us the displacement field which in turn yields the velocity field (Mohayaee et al. 2005). Here we use a catalog of 24,000 galaxies within a radius of  $8,000 \text{ km s}^{-1}$  we call ‘Nearby Galaxies –  $8,000 \text{ km s}^{-1}$ ’, or NBG–8k (provided by Brent Tully and not yet public). We present in Fig. 1a the reconstructed velocity field for this catalog.



**Fig. 1.** (a) Left panel: Reconstructed velocity field using the NBG–8k catalog. The slice is 20 Mpc/h deep. (b) Right panel: Reconstructed primordial density field using the NBG–8k catalog. Finger-of-God effect is only partially corrected for.

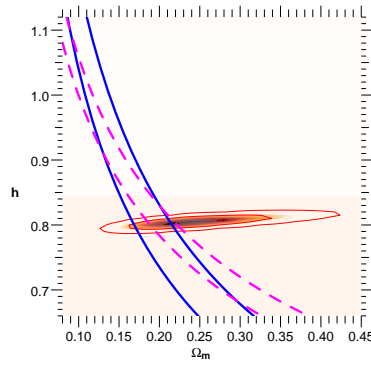
### 4.2 Primordial density field

The primordial density field is linked to the linear displacement field. The displacement field we are reconstructing is non-linear but only to the second Lagrangian order (as with the Zel’dovich approximation). Thus, we can use the relationship between Zel’dovich displacement and the density field to obtain the primordial density field from the non-linear displacement field. The result is shown in Fig. 1b.

### 4.3 $\Omega_m$ measurement

One of the applications of our method is a measurement of the local mean density of the universe. We may use the rules established in the former sections to build a correct set of initial data for the reconstruction. Then the likelihood function takes into account the special characteristics of the data in the catalog and the reconstructed velocity field. First, the error on distance modulus, and hence the absolute magnitude, is Gaussian distributed

(e.g. Pizagno et al. 2007). Second, the reconstruction introduces a Lorentzian distributed error on the velocity field. The result of the measurement assuming these two distributions is shown on Fig. 2. We see that the result is in perfect agreement with other measurements from the SDSS (Tegmark et al. 2004) and WMAP3. Spergel et al. (2007) gives the measurement for WMAP3+SDSS:  $\Omega_m = 0.266^{+0.026}_{-0.036}$ . Our likelihood function is maximal at  $(h, \Omega_m) = (0.805^{+0.005}_{-0.005}, 0.237^{+0.063}_{-0.056})$  at 68% of confidence level. The major axis of the ellipsoidal shape of the likelihood function can be fitted and we get  $h = 0.81 \left(\frac{\Omega_m}{0.3}\right)^{0.022}$ . This formula can be compared to WMAP's relation:  $\Omega_m h^2 = 0.127$  and SDSS's:  $h = 0.7 \left(\frac{\Omega_m}{0.3}\right)^{-0.35}$ .



**Fig. 2.** Measurement of  $\Omega_m$  based on NBG-3k catalog. The thin red lines are 68% and 95% of confidence limit for the measurement based on the catalog. The thick blue lines is coming from the WMAP3 experiment. The dashed magenta line is obtained by the SDSS project.

## 5 Conclusion

We have described how to use the MAK algorithm to accurately reconstruct the primordial density fluctuations above a few Mpc. The peculiar velocity field is accurately reconstructed provided one makes good assumptions on the mass-to-light ratio and takes into account the systematic errors. From it, we made a new apparently unbiased measurement of  $\Omega_m$  in agreement with WMAP and SDSS results.

This method is better than a previous approach like POTENT (Bertschinger & Dekel 1989). As it uses Lagrangian approximations, it handles the first Eulerian non-linearities better. Studies on mock catalogs prove that our method works optimally with respect to the Zel'dovich approximation. Moreover, it only needs redshift positions to reconstruct the peculiar velocity field, contrary to POTENT.

As for the future, large statistical uncertainties remain and further work is needed to reduce them, e.g. by using bigger catalogs to have better constraints on the Lagrangian volume. We may also want to improve the reconstruction by dropping the hypothesis of convex potential on the displacement field and including higher order gravity effects along the trajectories.

Once the primordial density field is reconstructed we will use it to resimulate the local universe and determine if we recover the structures which are present in the catalog.

## References

- Bertschinger, E., Dekel, A. 1989, ApL, 336, 5
- Bertsekas, D.P. 1988, Annals of Operations Research, Vol. 14, pp. 105-123
- Brenier, Y., Frisch, U., Hénon, M., Loeper, G., Mattarese, S., Mohayaee, R., Sobolevskii, A. 2003, MNRAS, 356, 501
- Mohayaee, R., Mathis, H., Colombi, S., Silk, J. 2006, MNRAS, 335, 939
- Peebles, P.J.E. 1989, ApJ, 344, 53
- Pizagno, J., Prada, F., Weinberg, D.H., et al., submitted to AJ, astro-ph/0608472
- Spergel, D.N., Bean R., Doré O., et al., submitted to ApJ, astro-ph/0603449
- Tegmark, M., Strauss, M.A., Blanton, M.R., et al. 2004, Phys. Rev. D, 69, 103501



---

## Using Local Volume data to constrain Dark Matter dynamics

G. Lavaux<sup>1</sup>, R. Mohayaee<sup>1</sup>, S. Colombi<sup>1</sup>, and R. B. Tully<sup>2</sup>

<sup>1</sup> Université Paris 6/Institut d'Astrophysique de Paris, 98bis Boulevard Arago,  
75014 PARIS, FRANCE [lavaux@iap.fr](mailto:lavaux@iap.fr)

<sup>2</sup> Institute for Astronomy, Univ. of Hawaii, Honolulu, USA

### 1 Introduction

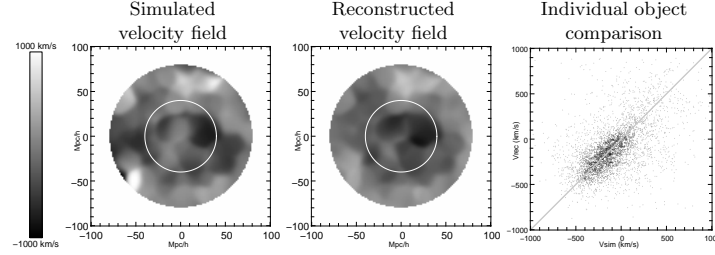
The peculiar velocities of galaxies hold a lot of information on the current dynamical state of the mass tracers such as galaxies or groups of galaxies in the Universe. This information can be used to constrain the relationship between the mass tracers and the underlying dark matter mass distribution responsible of the observed dynamic. This gives a measurement of  $\beta$ , the linear growth factor of density fluctuations, which yields either a measure of  $\Omega_m$ , the mean matter density in the Universe, or of the shape of the function  $M/L(L)$ . We propose here to show that Lagrangian reconstructions of peculiar velocities permit this kind of comparison and may serve to constrain the above two quantities. However these constraints can only be tight if observational biases are taken into account correctly, which is in most cases possible. Galaxies of the Local Volume, with good peculiar velocity measurements, may thus help constraining the dark matter dynamics.

Lagrangian reconstruction predicts peculiar velocities of individual mass tracers from only their redshift position and their mass. We will illustrate the reconstruction procedure by the means of the Monge-Ampère-Kantorovitch (MAK) reconstruction, which has already been extensively tested on simulations (Brenier et al., 2003; Mohayaee et al., 2006). The impact of observational biases such as redshift distortion, mass-to-light assignment (diffuse mass,  $M/L$  relation), incompleteness effect, finite volume effects, statistical measurement of  $\beta$  has only been studied thoroughly in Lavaux et al. (2007). In this proceeding, we propose to look more precisely at the consequence of choosing a specific mass-to-light assignment on detected tracers.

To study the aforementioned effects, we use a  $128^3$  N-body collisionless sample in a  $200^3 h^{-3}\text{Mpc}^3$  volume (Mohayaee et al., 2006), with  $\Omega_m = 0.30$ ,  $\Omega_\Lambda = 0.70$ . From a halo catalogue, *FullMock*, built from the sample, we have produced mock catalogues. *FullMock* is obtained using a standard Friend-of-friend algorithm, whose linking parameter is chosen to be a fifth of the mean particle separation of the original sample. Haloes of less than 5 particles were

left unbound and the corresponding particles put in a set called the “background field”. This set represents 63% of the total mass of the simulation.

## 2 Correctable observational biases



**Fig. 1.** Result of a reconstruction including finite volume effects (selection and boundary effects), redshift distortion, incompleteness. Left panel: simulated velocity field smoothed with a  $5 h^{-1}\text{Mpc}$  Gaussian filter. The white circle shows the  $40 h^{-1}\text{Mpc}$  region inside the  $80 h^{-1}\text{Mpc}$  deep mock catalogue. Middle panel: reconstructed velocity field smoothed equally. Right panel: individual comparison between reconstructed and simulated velocities for haloes of the mock catalogue inside the white circle.

The result of including all the observational effects mentioned in the introduction, except the mass-to-light assignment problem, in mock catalogues are given in Fig. 1. As one can see for a standard mock catalogue, built to have approximately the same features as observed redshift catalogues (such as 2MASS Redshift Survey, 2MRS hereafter), there is no significant bias introduced on reconstructed velocities. However the scatter may be significant, all the more that the number of measurements is low, which means a good statistical description of the error is needed to achieve an unbiased Bayesian estimator of  $\Omega_m$  from measurements. A problem that is not highlighted in this figure, is that incompleteness correction, if done unwisely, *e.g.* by assuming a wrong  $\Omega_m$ , may yield an offset in the relation shown in the right panel of Fig. 1. Unfortunately, a similar offset is introduced by a wrong choice of the Hubble constant  $H$  and thus its measurement may be hindered by incompleteness effects though hopefully without strongly affecting the  $\Omega_m$  measurement.

## 3 Mass-to-light assignment

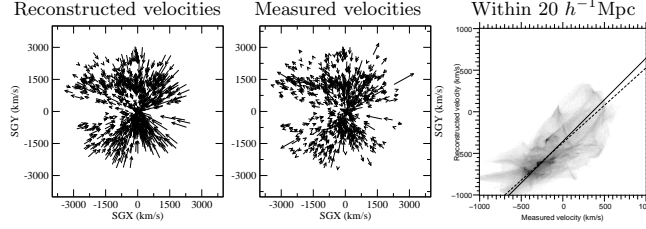
*Diffuse mass* – The first problem comes from the intrinsic limitation of redshift catalogues: they have a lower luminosity cut-off which translates to a

minimum lower mass. In fact, in 2MRS, with  $L_{\text{min,K}} \simeq 10^9 L_{\odot}$ , and using Sheth and Tormen (2002), we find that  $\simeq 50\%$  of the mass distribution may be missing in real data. We have tried to reproduce that problem in *FullMock* by not using the background field to build the reconstruction mesh. It has lead to an overestimation of the reconstructed peculiar velocities because the gravitational field of the haloes is in reality screened by the background field. On the contrary, if one predicts the amount of missing mass and introduces it homogeneously in the reconstruction mesh, reconstructed peculiar velocities are underestimated because of a too important screening effect. This two extreme cases lead, for mock catalogues, to low constraints on  $\Omega_m$ :  $0.15 \leq \Omega_m \leq 0.74$ . However we showed there is an optimal compromise between those two situations that give an unbiased reconstruction, with nearly no extra scatter on reconstructed velocities, when compared to a reconstruction on *FullMock*.

*M/L relation* – Another fundamental problem of mass-to-light assignment is the use of a unique  $M/L$  relation to transform detected galactic luminosities to dynamical masses. We have built a luminosity catalogue from *FullMock* using the  $M/L$  relation given by Tully (2005). A mass tracer catalogue is then recovered by assuming either a constant  $M/L$  or a  $M/L$  as given by Marinoni and Hudson (2002) which is obtained by mapping a Press-Schechter mass function to a Schechter luminosity function. The reconstructed velocities are then compared to the simulated ones. We tried also to introduce a significant uniform scatter on the logarithm of the mass  $M$  of haloes in the original mock catalogue. The width of the distribution was chosen with  $\Delta \log_{10} M = 1$  to be consistent with observations. The result of these three tests is that a random uncertainty on the  $M/L$  relation does not produce bias and leads only to a relatively small increase of the scatter between reconstructed and simulated velocities, whereas changing the relation (even a little) quickly introduces a significant systematic when comparing these same velocities.

#### 4 Application to an Extended Nearby Galaxies Catalog

While these results are still preliminary (Lavaux et al. in preparation), we present here a MAK reconstruction for which we have tried to account for redshift distortion, zone of avoidance, finite volume effect and incompleteness. We chose  $M/L = 300$  for elliptical galaxies and  $M/L = 100$  for spiral galaxies. This choice is motivated by observations of velocity dispersion of groups. As we have neglected the influence of the diffuse mass, we are most likely going to overestimate reconstructed peculiar velocities and thus to underestimate  $\Omega_m$ , as shown in § 3. Fig. 2 summarises the result obtained using this reconstruction. The reconstructed line-of-sight peculiar velocities look in agreement with measurements, at least in the central region. Of course, in the outer part of the catalogue measured velocities are both intrinsically noisy and incorrectly reconstructed because of boundary effects. The right panel shows



**Fig. 2.** Left panel: line-of-sight component of reconstructed velocities for objects with a measured velocity. Middle panel: same as left panel but for the measured velocities. Right panel: Comparison between reconstructed and measured velocity field, obtained after adaptive smoothing of the corresponding individual velocities on a uniform grid. Solid line corresponds to  $\Omega_m = 0.30$ , dashed-line to  $\Omega_m = 0.38$  (result given by the Bayesian estimator).

that reconstructed velocities are well correlated to measurements but that there is an offset most probably coming from an incompleteness correction defect, mentioned in § 2, that needs be corrected in future reconstructions.

## 5 Conclusion and perspective

So far we have checked the influence of each observational bias and shown what must be included to extract useful informations on dark matter dynamics. The first applications of our method to real galaxy catalogues look successful though are still probably affected by some of the biases. More checking must be conducted as has been done on mock catalogues to establish the amount of systematics. We are now going to apply the method to a bigger catalogue like 2MRS or 6dF to decrease finite volume effects. Other information may be extracted from Lagrangian reconstruction of the Local Volume such as local initial conditions for constrained simulations and statistical quantities (power spectrum, cosmological parameters) that may be compared to those obtained by CMB experiments.

## References

- Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskii. *MNRAS*, **346**, 501 (2003).  
R. Mohayaee, H. Mathis, S. Colombi, and J. Silk. *MNRAS*, **365**, 939 (2006).  
G. Lavaux, R. Mohayaee, S. Colombi, R. B. Tully, F. Bernardeau, and J. Silk. *ArXiv e-prints*, 707 (2007).  
R. K. Sheth and G. Tormen. *MNRAS*, **329**, 61 (2002).  
R. B. Tully. *ApJ*, **618**, 214 (2005).  
C. Marinoni and M. J. Hudson. *ApJ*, **569**, 101 (2002).



## **Annexe G Proceedings référés**

## Lagrangian reconstruction of cosmic velocity fields

Guilhem Lavaux<sup>1,2,3,\*</sup>

<sup>1</sup>*Institut d'Astrophysique de Paris, 98bis Bd Arago, 75015 PARIS, FRANCE*

<sup>2</sup>*Université Paris 6*

<sup>3</sup>*Université Paris 11*

We discuss a Lagrangian reconstruction method of the velocity field from galaxy redshift catalog that takes its root in the Euler equation. This results in a “functional” of the velocity field which must be minimized. This is helped by an algorithm solving the minimization of cost-flow problems. The results obtained by applying this method to cosmological problems are shown and boundary effects happening in real observational cases are then discussed. Finally, a statistical model of the errors made by the reconstruction method is proposed.

PACS numbers: 47.10.A-, 47.15.km, 47.11.Fg, 95.35.+d, 98.62.Py

### I. INTRODUCTION

Cosmologists are highly interested in studying galaxy peculiar velocities. Indeed, their study is a direct way to measure the dynamical state of a system and would thus permit to better understand dark matter distribution in our local Universe. The main difficulty is that measured velocities are only available sparsely and hence does not provide a good probe of the matter distribution. One must then devise an algorithm that is able to predict, under fair hypotheses, galaxy peculiar velocities from their present positions, which are their sky coordinates and their redshift, and compare the result to the measurement. Jim Peebles [1] first tried to do full orbit reconstruction by evolving the present system back in time. This method proved to be quite accurate for very small volume and number of objects. However, whenever one tries to reconstruct orbits of a large number of galaxies, the method fails because the number of plausible solution is blowing up. A simplification of this problem is presented: 3D galaxy positions are assumed to be known and a simpler gravitational dynamic model is going to be assumed. We will also assume that the dynamics of galaxies is mostly driven by collisionless dark matter particles.

This proceeding is organized as follows. In § II, we recall the principal result of the reconstruction method developed by [2] (see also the companion paper Mohayaee & Sobolevskii, hereafter MS). The method requests the using of a special fast algorithm to solve the problem. This algorithm is presented in § III. The method is then applied to a dark matter distribution obtained from a cosmological simulation and the reconstructed velocities are checked against the simulated ones (§ IV). Finally, a discussion on problems with bad boundary conditions, as usually met in observational cosmology, is quickly discussed in § V.

### II. VELOCITY RECONSTRUCTION THEORY

The theory of velocity reconstruction in cosmology is detailed by MS. We recall here the main results. To reconstruct the peculiar velocity field one must first compute the displacement field of dark matter particles by solving a Monge-Ampère equation [Eq.(16) of MS]. We achieve that by minimizing Eq.(17) of MS in its simplified form using the “Auction” algorithm, with  $\sigma$  the pairing map and  $\mu$  the mass of each particles of the mesh:

$$S_\sigma = \mu \sum_{i=0}^N (\mathbf{x}_i - \mathbf{q}_{\sigma(i)})^2, \quad (1)$$

The minimization is conducted over  $\sigma$ . We recall also the Zel’dovich approximation [Eq.(12) of MS] for the velocity field is taking the following form

$$\mathbf{v}(\mathbf{x}_i) = \beta (\mathbf{x}_i - \mathbf{q}_i) \quad (2)$$

where  $\beta$  is the linear growth factor, which is well approximated by  $\beta \simeq \Omega_m^{9/5}$  when it is computed at redshift  $z = 0$ .

### III. MINIMIZATION ALGORITHM

Direct minimization of Eq. (1) is a computationally difficult problem [time complexity  $O(N!)$ ]. Fortunately, there exist better alternatives that have been developed for solving minimal cost flow problems which can be adapted to our minimal transportation problem. In particular, we are going to use the “Auction” algorithm developed in [3]. The time complexity of this algorithm is of the order of  $O(n^{2.25})$  by direct performance measurement, with  $n$  the particle density [7]. The exact constant hidden in  $O(n^{2.25})$  depends a lot on the difficulty of the assignment problem, which means it is catalog dependent.

\*Electronic address:  
URL: <http://www.iap.fr/users/lavaux/>

lavaux@iap.fr;

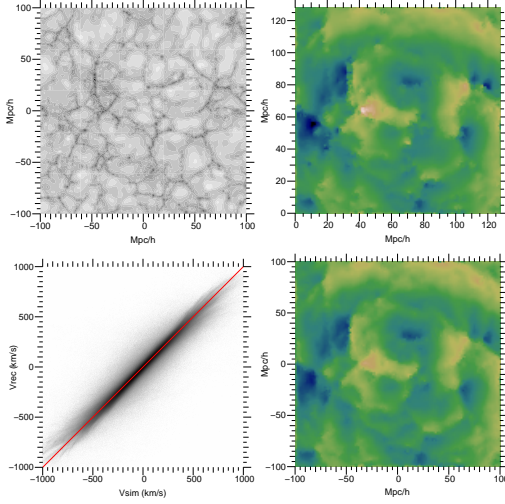


FIG. 1: *Application to Cosmology* – *Top left*: A slice of the density field of the  $\Lambda$ CDM simulation that is used for the tests (shades of gray indicates logarithm of the mass density). *Top right*: Adaptively smoothed line-of-sight component of the velocity field in the same slice. *Bottom right*: MAK reconstructed line-of-sight component of the velocity field of the same slice. Linear color scale: dark blue =  $-1000 \text{ km s}^{-1}$ , white =  $+1000 \text{ km s}^{-1}$ . *Bottom left*: Scatter plot between reconstructed and simulated velocities for objects identified in the simulation. Shades of grey show levels of the logarithm of the point density.

### A. Auction algorithm

The algorithm tries to evolve the pairing map  $\sigma$  between  $\mathbf{x}_i$  and  $\mathbf{q}_j$  such that when the function is stationary between two consequent iteration it corresponds to minimizing the given total association cost. Particles locate at different Eulerian positions  $x_i$  compete against each other for Lagrangian positions  $q_j$ . Minimization of the total association cost  $S_\sigma$  is achieved by studying the dual problem of minimization of association penalties  $p_j$ . In [3] it is shown that

$$\min_{\sigma} S_{\sigma} = \max_{p_j: j=1, \dots, n} \left\{ \sum_j p_j + \sum_i r_i \right\}, \quad (3)$$

with  $a_{i,j} = \mu(\mathbf{x}_i - \mathbf{q}_j)^2$ , the cost of associating  $\mathbf{x}_i$  to  $\mathbf{q}_j$  and  $r_i = \min_j (a_{i,j} + p_j)$ . Once the set  $\{p_j\}$  is determined by the above maximization, the map  $\sigma$  is simply given by:

$$\sigma(i) = \arg \min_j \{a_{i,j} + p_j\}. \quad (4)$$

Effectively,  $\{p_j\}$  is computed iteratively by the algorithm. Each iteration is composed of two parts. During the first

one, we obtain a set of best assignment  $\mathcal{A}(j)$  for each particle  $\mathbf{q}_j$  by minimizing all possible  $r_i$ . Then, we link  $\mathbf{x}_{i_j^*}$  to  $\mathbf{q}_j$  with  $i_j^*$  being the particle having the minimal  $r_{i_j^*}$  in the set  $\mathcal{A}(j)$ . We also have a reverse mapping for this link that we write  $j_i^*$ . Finally, the penalty  $p_j$  is updated such that

$$p_j \rightarrow \tilde{p}_j = a_{i_j^*,j} + w_{i_j^*} - \epsilon, \quad (5)$$

with  $\epsilon > 0$  and

$$w_i = \min_{j \neq j_i^*} (a_{i,j} + p_j). \quad (6)$$

The solution found is the same as for  $\epsilon = 0$  provided  $\epsilon < \epsilon_0/N$ , with

$$\epsilon_0 = \min_{\{i,j\}/a_{i,j} \neq 0} a_{i,j}. \quad (7)$$

The time complexity depends quite a lot on the way  $\epsilon$  is scaled down from its initial value to the  $\epsilon_0/N$ . Numerical experiments have shown that trying to converge in about 5 iterations and starting from  $\epsilon/\epsilon_0 \simeq N/2$  seems to give a faster convergence.

### B. Implementation

We developed a C++ multithreaded (shared memory parallelism) and MPI version of the “Auction” algorithm, it will be available later as a multi-purpose library for cost-flow problems at the address <http://www.iap.fr/users/lavaux/>. Besides doing a full minimization over all  $\mathbf{q}_j$  for a given  $\mathbf{x}_i$  (“dense” mode). It also supports a “sparse” mode that solves a partial minimization problem: for a given  $\mathbf{x}_i$ , it only minimizes over a subset of  $\{\mathbf{q}_j\}$  such that  $\|\mathbf{x}_i - \mathbf{q}_j\|_{\infty} < R$ , where  $R$  is a parameter given at the initialization to the algorithm. This allows to reduce drastically the computing time while giving the same result provided that  $R$  is not too small (typically  $R = 40 h^{-1} \text{ Mpc}$  for a  $\Lambda$ CDM Universe). On a Dual-core AMD Athlon64 4800+, the SMP implementation (dense mode) takes 50 mins to assigning 79,000 particles. It has successfully reconstructed a  $128^3$  dense mesh in a month in sparse mode. The MPI version of the corresponding algorithm is only performant for larger number of particles (typically  $N \gtrsim 500,000$ ). Most of the time is, at the moment, spent at computing  $\min_j (a_{i,j} + p_j)$  as the cost values are only kept in a minimalistic cache. Precomputing the costs is also not feasible because of the excessive amount of memory that would be needed to store all costs for all  $(i, j)$  pairs. We also consider to implement a general purpose totally asynchronous implementation in the near future.

### IV. APPLICATION TO COSMOLOGY: TEST ON COSMOLOGICAL SIMULATION

To check that the dark matter dynamical model is working, we are testing it against a  $128^3$   $N$ -body sample



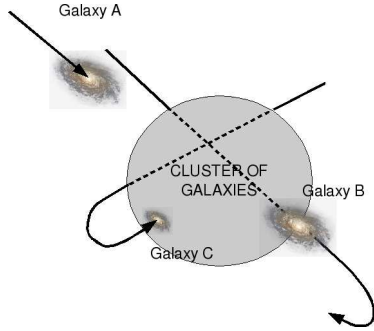


FIG. 2: Cosmology / Multi-streaming regions – This figure illustrates the different problems that may occur for a halo of dark matter particles near a cluster of galaxies. Galaxy A is in the region of first infall. The displacement field will be well reconstructed. Galaxy B is coming from the same direction as Galaxy A but has already gone through the center of the cluster and is decelerating. In that case, its displacement is badly reconstructed as, most likely, MAK predicts that the matter composing Galaxy B is coming from the region opposite to Galaxy A's region. Galaxy C is also wrongly reconstructed.

[4] which was generated with the public version of the  $N$ -body code HYDRA [5] to simulate collisionless structure formation in a standard  $\Lambda$ CDM cosmology. The volume of the simulation is  $200^3 h^{-3} \text{ Mpc}^3$ . The mean matter density is  $\Omega_m = 0.30$  and the cosmological constant is  $\Omega_\Lambda = 0.70$ . The Hubble constant is  $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and the normalization of the density fluctuations in a sphere of radius  $8 h^{-1} \text{ Mpc}$  is  $\sigma_8 = 0.99$ .

Halo of dark matter particles are identified using a friend-of-friend algorithm with a traditional value of the linking parameter  $l = 0.2$  times the mean particle separation. A limit of 5 linked particles is put to bind particles into a halo. The particles left unbound by this criterion were kept in a set called the “background field”. All objects are kept in a mock catalogue called *FullMock*. We have run a reconstruction on *FullMock* using a MAK mesh with  $128^3$  elements. Each object of *FullMock* was given a number of elements  $\mathbf{x}_i$  equal to the number of particles of the original simulation which has been bound into this object. We distributed the  $\mathbf{q}_j$  mesh elements regularly on a cubic grid of the same physical size as the simulation box. Finally we computed the convex mapping  $\sigma$  corresponding to the MAK problem with the help of the algorithm described in § III. The velocities for each particle were computed using the Zel'dovich approximation Eq. (2), using the same cosmology as the simulation to compute  $\dot{D}(t)$ .

Fig. 1 summarizes the results obtained using the MAK method on the reconstructed velocities. The individual object velocities, in the bottom-left panel, are exceptionally well reconstructed. Visual inspection of the line-

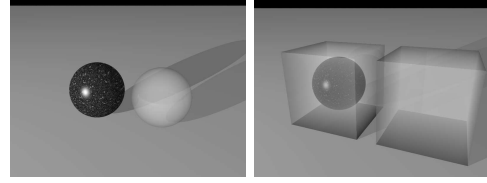


FIG. 3: Cosmology / Boundary problems – Left panel: Illustration of the NaiveDom approach to handle boundary problems while doing a reconstruction. The dark starry ball illustrates the current dark matter distribution as inferred from galaxy catalogues. The whitish transparent ball is the assumed initial volume for the dark matter that has fallen in present structures. Right panel: Same as left panel but this illustrates the PaddedDom approach.

of-sight component of the velocity field in the two right panels show nearly no discrepancy except in regions with really high velocities. In these regions, the dynamics is highly non-linear, which means that the convex hypothesis is not valid anymore. This problem arises on a typical cosmological scale of at most a few Mpc around large clusters. Indeed, in those regions the fluid description of dark matter particles completely fail because the mass tracers may have already crossed the center of the gravitational attractor and are currently falling back to the center, as illustrated Fig. 2. This renders the displacement field reconstruction dubious in those cases.

## V. APPLICATION TO COSMOLOGY: BOUNDARY PROBLEMS

One does not necessarily know the Lagrangian domain  $\mathbf{q}$  on which the MAK reconstruction must be computed. This is the case for real cosmological observations and one must use some empirical prescription to attenuate the boundary effects on reconstructed velocities. This scheme is helped by the overall homogeneity of the Universe above scales larger than  $200 h^{-1} \text{ Mpc}$ . We propose thus to check two schemes to handle boundary effects:

- A naive approach would be to assume that the piece of Universe considered has not changed its volume sufficiently between initial time and the current time. This means that we may assume that if we select a ball of matter, in the Universe, centered on us, all the mass that is inside this ball is coming from the same homogeneous ball in the Universe as it was at decoupling time. We call this approach *NaiveDom*. It is equivalent to say that tidal field effects are totally negligible on the considered scale.
- An alternative approach is not to make an assumption on the exact shape but on the low amount of fluctuation on the boundary. Consequently, if one

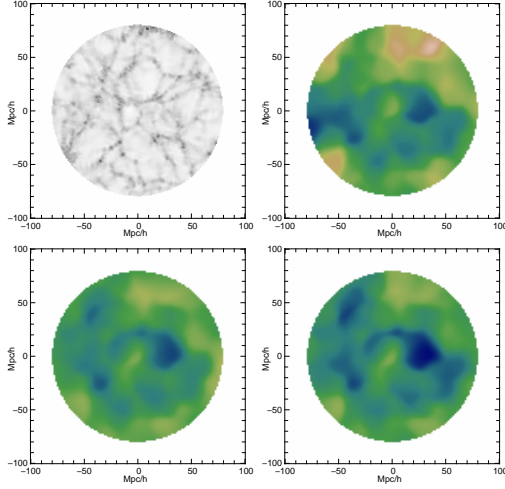


FIG. 4: *Cosmology / Boundary problems* – Outer boundary problems while doing reconstruction on finite volume catalogue. Color scale is the same everywhere (dark blue=−1000 km/s, white=+1000 km/s). *Top left*: Density field of the mock catalogue (log scale). *Top right*: Simulated velocity field, smoothed with a  $5 h^{-1}$  Mpc Gaussian window. *Low left*: PaddedDom velocity field, smoothed equally. *Low right*: NaiveDom velocity field, smoothed equally.

selects the same ball of matter in the present Universe, it is fair under this approximation to pad the matter distribution using homogeneously distributed particles. One may then build the mapping between the “padded piece of Universe” and an initial completely homogeneous set of particles. We call this approach *PaddedDom*.

These two ways of handling boundary effects are illustrated Fig. 3 and the results are presented in Fig. 4.

As expected, boundaries are badly reconstructed in PaddedDom and NaiveDom. However at the center of the spherical cut, the velocity field seems correctly reconstructed by visual comparison to the velocity field computed from the simulation. Looking carefully at the result using NaiveDom indicates that there is likely a systematic error near the center (the blue region is darker and more extended than in the two other figures). This is probably due to stronger boundary effects that are not correctly attenuated by the NaiveDom scheme (a detailed quantitative analysis of boundary artefacts are given in [6]). Empirically, we found that a buffer zone of, at least, about  $20 h^{-1}$  Mpc is needed to reduce boundary effects with a PaddedDom reconstruction scheme.

## VI. STATISTICAL ANALYSIS OF ERRORS IN THE RECONSTRUCTION

The measurement of the slope between velocities and reconstructed displacements should give an estimation of  $\Omega_m$ . However, building a reliable estimator of this slope without the statistical model of errors made both at the observation and the reconstruction level may produce unacceptable bias. We propose to show how to use models on reconstruction errors to make a bayesian analysis of the reconstructed velocities. We will focus here on errors made during a reconstruction and assume that the observed peculiar velocities  $v$  are equal to their true velocities. A more detailed discussion can be found in [6].

Using simulations, we have measured the distribution of reconstruction errors, for each object  $i$  of a catalog of galaxy,  $\{e_i\}$  defined as

$$e = v_r - \beta \psi_{r, \text{rec}}, \quad (8)$$

with  $\beta = 0.51$  for the studied simulation (corresponding to  $\Omega_m = 0.30$ ),  $v_r$  the line-of-sight component of the simulated velocity of the considered,  $\psi_{r, \text{rec}}$  the reconstructed radial displacement. The result is given in Fig. 5. We have tried to fit an histogram of the errors  $\{e_i\}$  by both a Gaussian function of width  $B$

$$f_G(e) \propto \exp\left(-\frac{e^2}{B^2}\right) \quad (9)$$

and a Lorentzian function

$$f_L(e) \propto \frac{1}{1 + \frac{e^2}{B^2}}. \quad (10)$$

We obtained approximately the same width  $B$  for the two fits (which is expected from the second order development of both functions), however it is striking that  $f_L$  is a much better approximation than  $f_G$  to the observed error distribution.

We equate the probability of getting an error  $e$  on the true velocity  $v_r$  for an object of the catalog to  $f_L(e)$ . We also assume now that the distribution of velocities in the object sample is, for a sufficiently large volume, Gaussian with a width  $\sigma_v$ :

$$P(v_r | \sigma_v) \propto \exp\left(-\frac{v_r^2}{2\sigma_v^2}\right). \quad (11)$$

Now we can build the joint probability of getting  $v_r$ ,  $\psi_{r, \text{rec}}$  and  $\beta$ :

$$\begin{aligned} &P(v_r, \psi_{r, \text{rec}}, \beta | B, \sigma_v) \\ &\propto P(e(v_r, \psi_{r, \text{rec}}) | B, \sigma_v) P(v_r | B, \sigma_v) P(\psi_{r, \text{rec}} | B, \sigma_v) \\ &\propto P(\psi_{r, \text{rec}} | B, \sigma_v) \frac{\exp\left(-\frac{v_r^2}{2\sigma_v^2}\right)}{1 + \left(\frac{v_r - \beta \psi_{r, \text{rec}}}{B}\right)^2}, \end{aligned} \quad (12)$$

where the constant of proportionality eventually depends on  $B$ ,  $\sigma_v$  and  $\beta$ . Using the theorem of Bayes, it is now

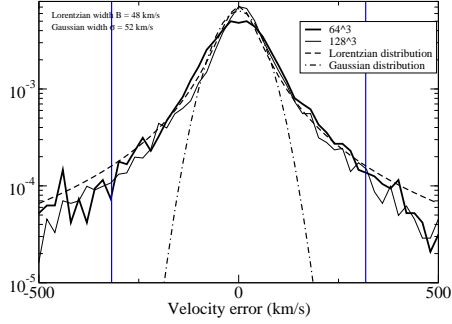


FIG. 5: Error in the reconstruction – This plot displays the probability distribution of the quantity  $v_{r,\text{rec}} - v_{r,\text{sim}}$ , where  $v_{r,\text{rec}}$  and  $v_{r,\text{sim}}$  are the line-of-sight reconstructed and simulated velocities, respectively, after choosing an observer at the center of the simulation box. The dashed and dot-dashed curve give the best fit of a Gaussian and a Lorentzian distribution, respectively.

possible to compute the conditional probability that the true velocity of some object is  $v_r$  given that the reconstructed displacement is  $\psi_{r,\text{rec}}$ :

$$P(v_r | \psi_{r,\text{rec}}, \beta, B, \sigma_v) = \frac{e^{-\frac{v_r^2}{2\sigma_v^2}} \left(1 + \left(\frac{\beta_* \psi_{r,\text{rec}} - \alpha_* v_r + \gamma_*}{B_v}\right)^2\right)^{-1}}{\int_{v=-\infty}^{+\infty} e^{-\frac{v^2}{2\sigma_v^2}} \left(1 + \left(\frac{\beta_* \psi_{r,\text{rec}} - \alpha_* v + \gamma_*}{B_v}\right)^2\right)^{-1} dv} \quad (13)$$

To obtain the total likelihood  $\mathfrak{L}(\beta)$  to observe true velocities  $\{v_{i,r}\}$  given that the reconstructed displacements are  $\{\psi_{i,r,\text{rec}}\}$ , one may assume the statistical independence of the  $(v_{i,r}, \psi_{i,r,\text{rec}})$  duets. With this assumption,  $\mathfrak{L}$  is simply

$$\mathfrak{L}(\beta) = \prod_i P(v_{i,r} | \psi_{i,r,\text{rec}}, \beta, B, \sigma_v) \quad (14)$$

Using that approach we have made measurements in finite volume mock catalogs. For example, with a Padded-Dom reconstruction, one measure  $\Omega_m = 0.34$  with this approach (for an effective  $\Omega_m = 0.35$  in this catalog), whereas a naive measurement would yield  $\Omega_m \leq 0.26$ .

## VII. CONCLUSION

We presented a method to predict velocities of galaxies from their current position. To solve this problem, we implemented a fast algorithm invented by Dimitri Bertsekas [3] and applied the method to a pure dark matter simulation. It happens that the reconstructed velocities are impressively accurate on large-scales (§ IV). However, the solution is only approximate in regions where multi-streaming occurs.

We proposed two methods for partially correcting boundary effects (§ V) and showed how boundary effects affect the reconstructed velocity field. We preferred the PaddedDom reconstruction scheme as it seems to give overall better results. Empirically we found that a buffer zone of  $20 h^{-1}$  Mpc is needed before obtaining a reconstructed velocity field correlated with the one given by the simulation.

At last, we proposed a bayesian model (§ VI) to account for reconstruction errors while estimating the slope between the reconstructed displacements and the true velocities of objects in a galaxy catalogs.

We would like to continue this work by improving the padding schemes to have even less boundary effects and make full use of available data in astronomy. We are also working on an improved algorithm that is able to take into account in a better way the non-linearities that are introduced in the velocity field due to gravitational effects occurring along particle trajectories. This new algorithm will try to fully solve the Euler-Poisson problem.[8]

This work is partially supported by the ANR grant BLAN07-2\_183172 (project OTARIE).

- [1] P. J. E. Peebles, Tracing galaxy orbits back in time, *ApJL* 344 (1989) L53–L56.
- [2] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, A. Sobolevskii, Reconstruction of the early Universe as a convex optimization problem, *MNRAS* 346 (2003) 501–524.
- [3] D. P. Bertsekas, *A Distributed Algorithm for the Assignment Problem*, MIT Press, Cambridge, MA, 1979.
- [4] R. Mohayaee, H. Mathis, S. Colombi, J. Silk, Reconstruction of primordial density fields, *MNRAS* 365 (2006) 939–959.
- [5] H. M. P. Couchman, P. A. Thomas, F. R. Pearce, Hydra:

an Adaptive-Mesh Implementation of P 3M-SPH, *ApJ* 452 (1995) 797–+.

- [6] G. Lavaux, R. Mohayaee, S. Colombi, R. B. Tully, F. Bernardreau, J. Silk, Observational biases in Lagrangian reconstructions of cosmic velocity fields, *ArXiv e-prints* 707.
- [7] This number is obtained for a given simulation, and particles randomly until the desired average density is obtained. The worst case of this algorithm is actually  $O(N^3)$ , if one makes a dense search on purely random data.
- [8] G. Lavaux & G. Loeper, work in progress.

## **Annexe H Article référé**

## Observational biases in Lagrangian reconstructions of cosmic velocity fields

G. Lavaux<sup>1</sup>, R. Mohayaee<sup>1</sup>, S. Colombi<sup>1</sup>, R. B. Tully<sup>2</sup>, F. Bernardeau<sup>3</sup>, J. Silk<sup>1,4</sup>

<sup>1</sup> Institut d'Astrophysique de Paris – UMR 7095, 98bis bd Arago, France, CNRS/Université Pierre et Marie Curie

<sup>2</sup> Institute for Astronomy, Univ. of Hawaii, Honolulu, USA

<sup>3</sup> Service de Physique Théorique, CEA/DSM/SPHT, Unité de recherche associée au CNRS, CEA/Saclay 91191 Gif-sur-Yvette cedex

<sup>4</sup> Department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH

14 September 2008

### ABSTRACT

Lagrangian reconstruction of large-scale peculiar velocity fields can be strongly affected by observational biases. We develop a thorough analysis of these systematic effects by relying on specially selected mock catalogues. For the purpose of this paper, we use the Monge-Ampère-Kantorovitch (MAK) reconstruction method, although any other Lagrangian reconstruction method should be sensitive to the same problems. We extensively study the uncertainty in the mass-to-light assignment due to incompleteness (missing luminous mass tracers), and the poorly-determined relation between mass and luminosity. The impact of redshift distortion corrections is analyzed in the context of MAK and we check the importance of edge and finite-volume effects on the reconstructed velocities. Using three mock catalogues with different average densities, we also study the effect of cosmic variance. In particular, one of them presents the same global features as found in observational catalogues that extend to  $80\ h^{-1}\text{Mpc}$  scales. We give recipes, checked using the aforementioned mock catalogues, to handle these particular observational effects, after having introduced them into the mock catalogues so as to quantitatively mimic the most densely sampled currently available galaxy catalogue of the nearby universe. Once biases have been taken care of, the typical resulting error in reconstructed velocities is typically about a *quarter* of the overall velocity dispersion, and without significant bias. We finally model our reconstruction errors to propose an improved Bayesian approach to measure  $\Omega_m$  in an unbiased way by comparing the reconstructed velocities to the measured ones in distance space, even though they may be plagued by large errors. We show that, in the context of observational data, it is possible to build a nearly unbiased estimator of  $\Omega_m$  using MAK reconstruction.

**Key words:** dark matter — cosmological parameters — methods: analytical and numerical — galaxies: distances and redshifts

### INTRODUCTION

Galaxy redshift catalogues provide us with the radial velocities of the galaxies,

$$cz = H_0 r + v_r, \quad (1)$$

which are partly due to the global Hubble expansion ( $H_0 r$  with  $H_0$  the present value of the Hubble parameter) and partly due to the line-of-sight components of the peculiar velocities ( $v_r$ ). Peculiar velocities are the deviations of galaxy velocities from the uniform Hubble expansion, due to the non-homogeneous distribution of matter in the Universe. The peculiar velocities are thus tracers of mass distribution in the Universe and can have far-reaching implications for cosmology. As tracers of dark matter, peculiar veloci-

ties can be used to determine the local and global distribution of dark matter. From expression (1), it is evident that observations of galaxy redshifts ( $z$ ) supplemented by measure of radial distances ( $r$ ), would yield the peculiar velocities. However, measuring distances is a non-trivial exercise. The Tully-Fisher relation, surface brightness fluctuations, the Faber-Jackson relation for ellipticals (and their siblings, including the fundamental plane and the  $D_n - \sigma$  methods, the Tip of the Red Giant Branch, Cepheids, and SNIa are the most usual methods for obtaining distances. The data gathered is however rather sparse: out of about a million galaxies whose redshifts are presently known with surveys such as 2dF and SDSS, the distances to only a few thousand have measured distances. Moreover, distances for most

of these galaxies have too large peculiar velocity errors (due essentially to errors in distance measurements) to be useful in studying dynamics. For instance, distance indicators such as the Tully-Fisher relation suffer from 20% relative distance errors and thus produce quite noisy measurements at relatively moderate redshifts (*i.e.*  $cz \gtrsim 3000 \text{ km s}^{-1}$ ). The data also suffers from selection biases (Strauss & Willick 1995; Tully & Pierce 2000). One way of reducing the error bars on distances is to average over many distance measurements for galaxies in clusters or groups and also by combining the results from different distance estimators. This treatment decreases the error bars on distances to about 8% relative distance errors (Tully et al. 2007). Even though all these difficulties can be surmounted, one can finally hope to only have a sparse sample (as compared to redshift samples) of radial components of peculiar velocities. Fortunately, we now have Lagrangian velocity reconstruction schemes that are based solely on current redshift positions of mass tracers. The reconstructed velocities depend on cosmological parameters. Thus, comparing predictions obtained through Lagrangian reconstruction algorithms and the measured velocities may give estimations of these parameters.

This brings us to the *main point* that this paper tries to address: developing a *robust* and *unbiased* method of Lagrangian peculiar velocity reconstruction using redshift catalogues, in particular when observational effects distort most of the required data needed for the reconstruction of the dynamics. The reconstructed velocities are then compared to the measured ones using an *ad hoc* algorithm to yield a measurement of  $\Omega_m$ , the mean matter density of the Universe.

Throughout the paper, we will try to mimic observational effects as they appear in the most densely sampled currently available galaxy catalogue of the nearby universe which has been compiled by one of the authors (R. B. Tully). This galaxy catalogue is built from different sources such as ZCAT (Huchra et al. 1992) and SSRS (da Costa et al. 1988). Only galaxies for which  $cz \leq 8000 \text{ km s}^{-1}$  have been introduced in the catalogue. This catalogue is named NBG-8k, standing for NearBy Galaxy catalogue with a depth of  $8000 \text{ km s}^{-1}$ . Although selection criteria for this catalogue are not well defined, it will prove to be useful for the study of smaller galaxy catalogues such as NBG-3k (Tully et al. 2007).

For the purpose of this paper, we use a recently developed technique, called the Monge-Ampère-Kantorovitch reconstruction method (MAK hereafter), which is an approximation to the full non-linear dynamics to trace orbits back in time. This is a Lagrangian method, such as PIZA (Croft & Gaztanaga 1997) or the Least-Action method (Peebles 1989), and not a Eulerian technique such as, *e.g.*, POTENT (Bertschinger & Dekel 1989). One must note that the results of this paper are also valid for the other Lagrangian reconstruction methods as all the effects we are going to analyze are explainable in terms of gravitational dynamics. The MAK reconstruction has already been largely discussed when applied on numerical simulations (Mohayaee et al. 2006; Brenier et al. 2003). It is based on assuming that the dark matter displacement field is convex and potential, *i.e.* irrotational. In doing so, we exclude displacement fields which include multistreaming regions. The main result is that it is then possible to reconstruct accurately and uniquely the displacement field of dark matter particles

between their original position and their current position. Practically, to solve the MAK problem, one must minimize a cost function for the assignment of a dark matter particle at the present comoving position  $\mathbf{x}_i$  and its initial comoving position  $\mathbf{q}_j$ :

$$S_\sigma = \sum_{i=1}^N (\mathbf{x}_i - \mathbf{q}_{\sigma(i)})^2. \quad (2)$$

If the Universe is assumed to be initially homogeneous, which is a fair hypothesis supported by CMB data (*e.g.* WMAP first year in Bennett et al. 2003),<sup>1</sup> then  $\mathbf{q}_j$  must be distributed on a uniform grid and the solution to the MAK problem is unique and given by the assignment  $\sigma$  which minimizes  $S_\sigma$ . The derived solution is then necessarily irrotational and derives from a convex potential. To solve this problem, we have implemented a parallel version of the so-called “auction” algorithm proposed by Bertsekas (1979).<sup>2</sup> Of course, as we are using an approximation to the dynamics, the solution to the problem will be only valid above some scale (typically a few  $h^{-1} \text{ Mpc}$ ). Once the solution is found, the immediate output of MAK reconstruction is the nonlinear displacement field  $\Psi(\mathbf{q}) = \mathbf{x}(\mathbf{q}) - \mathbf{q}$ , which can be used to find the peculiar velocity field  $\mathbf{v}$  using the first-order Zel’dovich approximation:

$$\mathbf{v}_i = \beta \Psi_i, \quad (3)$$

where the subscript  $i$  indicates the comparison is achieved on the corresponding field averaged over the object  $i$  (*i.e.* in a Lagrangian way), and the linear growth factor  $\beta \simeq \Omega_m^{5/9}$  (Bouchet et al. 1995). This best fit for  $\beta$  is valid as soon as  $\Omega_m + \Omega_\Lambda = 1$ ,  $\Omega_\Lambda$  being the present dark energy density. It appears then that a direct comparison of  $\Psi_i$  against  $\mathbf{v}_i$  should in principle give us  $\beta$  and thus  $\Omega_m$ . Though naive measurements (Mohayaee & Tully 2005) and preliminary studies (Branchini et al. 2002; Phelps et al. 2006) on mock redshift catalogues have already been tried, the observational biases and systematic errors in the velocity-velocity comparison have never been studied thoroughly.

This paper is organized as follows. In Section 1, we describe the simulation and the basic mock catalogues that are used in the rest of this paper. Subsequent mock catalogues integrate more and more observational features but are still based on the same original basic mock catalogues presented in this section. Section 2 gives a model for the error distribution on MAK velocities and discuss the first problematic features of the comparison between MAK and measured velocities. This error distribution helps us in particular to establish the likelihood analysis in Section 6. We

<sup>1</sup> Brenier et al. (2003) actually shows the uniformity is even required to prevent singularities in the solution of the Euler-Poisson system of equations.

<sup>2</sup> We implemented a parallel version for shared-memory supercomputers and MPI clusters. On the Magique2 cluster, it needs 50 minutes on 2 processors to solve the assignment of 74000 particles. The algorithm is already sparse, *i.e.* it only looks for candidates for assignment in a limited region of the catalogue. The MPI efficiency is here optimal using 2 processors. It must be noted that the time complexity depends highly on the catalogue that is being reconstructed. For a given catalogue, the time needed to solve the assignment problem increase as  $N^{2.25}$  with  $N$  the number of particles.



## Observational biases in Lagrangian reconstructions of cosmic velocity fields 3

go then to the first main topic of this paper in Section 3 by studying the systematic errors introduced by arbitrary mass-to-light assignments in redshift catalogues. This section includes a study of missing mass correction (§ 3.1), unknown  $M/L$  function (§ 3.2) and incompleteness effects (§ 3.3; technical details are given in Appendix C). In Section 4, we discuss the problem of redshift distortions and the way to account for it during the MAK reconstruction. Section 5 is devoted to the handling of finite volume and edge effects, i.e. issues related to the zone of avoidance (§ 5.1), the choice of the Lagrangian volume of the reconstruction (§ 5.2), and finally the so-called cosmic variance (§ 5.3). The last section (§ 6) of this paper investigates the effect of distance measurement errors on the comparison between reconstructed and measured velocities, and proposes a maximum likelihood estimator (§ 6.2) to account for them in the measurement of  $\Omega_m$ . Results given by this estimator are then discussed in § 6.3.

## 1 MOCK CATALOGUES

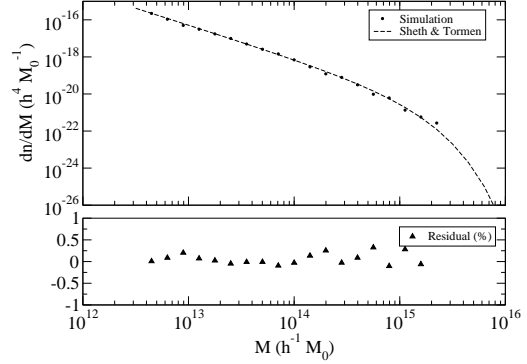
To study various effects and systematic biases on the MAK reconstructed velocity field, we generated a number of mock catalogues extracted from a  $N$ -body simulation (§ 1.1). Although many recipes will be employed later to address various observational biases, we will always start from the same three<sup>3</sup> “main” halo catalogues as described in § 1.2. The first catalogue aims to reproduce to some extent the main features of the local universe, in particular the presence of a large cluster at about  $40 h^{-1}\text{Mpc}$  and a super-cluster at about  $70 h^{-1}\text{Mpc}$ . The second and the third catalogues have less salient features but represent locally overdense and underdense realisations in order to address the problem of cosmic variance.

1.1 The  $N$ -body sample

Our  $128^3$  particles  $N$ -body sample (Mohayaee et al. 2006) was generated with the public version of the  $N$ -body code HYDRA (Couchman et al. 1995) to simulate collisionless structure formation in a standard  $\Lambda\text{CDM}$  cosmology. The sample covers a comoving volume of  $200^3 h^{-3} \text{Mpc}^3$ . The mean matter density is  $\Omega_m = 0.30$  and the cosmological constant  $\Omega_\Lambda = 0.70$ . The Hubble constant is  $H_0 = 65 \text{ km s}^{-1} \text{Mpc}^{-1}$ . The normalisation of the density fluctuations in a sphere of radius  $8 h^{-1} \text{Mpc}$ , is  $\sigma_8 = 0.99$ . We note that this value of  $\sigma_8$  is significantly larger than the value suggested by present WMAP data which sets  $\sigma_8 = 0.74$  (Spergel et al. 2006), but this should not affect significantly the results presented in this paper. In fact, a lower  $\sigma_8$  compared to 0.99 would reduce both non-linearities and cosmic variance effects, hence improving the quality of the measurements.

As the velocity field presents significant fluctuations on a larger scale than for the density field, one may worry about the small size of the simulation volume. We have checked, using linear theory, that the velocity dispersion in  $200^3 h^{-3} \text{Mpc}^3$ , for our cosmology, is  $40 \text{ km s}^{-1}$ . This value

<sup>3</sup> The computationally high cost of the reconstruction considerably limits the number of possible realisations.



**Figure 1.** *Sheth & Tormen mass function / Diffuse mass* – The top panel of this plot gives the number density of haloes in a mass bin as a function of the mass. The round points give the measurement of this function in the halo catalog whereas the dashed line is obtained using the Sheth & Tormen (2002) theory. The residuals between the prediction and the measurement are given in the lower panel (relative differences). Most of the time, the points are within a few percent of the theoretical prediction.

has to be compared to the typical errors appearing while doing velocity reconstructions to ensure that cosmic variance effects are negligible for our purpose.

## 1.2 The basic mock catalogues

To build mock catalogues, we have selected haloes from the  $N$ -body experiment using a standard Friend-Of-Friend algorithm with a traditional value of the linking parameter given by  $l = 0.2$  (Efstathiou et al. 1988). Haloes with less than 5 particles, i.e. with mass smaller than  $M_{\min} = 1.62 \times 10^{12} h^{-1} \text{M}_\odot$ , were discarded. Fig. 1 shows the good agreement between the measured halo mass function and the Sheth & Tormen (2002) model for haloes with  $M \gtrsim M_{\min}$ . However about 63% of the mass is not clumped in these haloes and is distributed in the *background field*. In realistic galaxy samples such as the NBG-8k or the 2MASS catalogue the lower mass cut-off is of the order of  $10^{11} \text{M}_\odot$ , a value much smaller than our  $M_{\min}$ . To mimic galaxies with mass smaller than  $M_{\min}$ , as will be required in the following, we just use dark matter particles unassigned to any halo as tracers. The catalogue containing all the haloes and all the field particles will be called *FullMock*. One could here worry that the  $N$ -body sample that we are using has a too low resolution as the spatial distribution of small halos is biased but not the particles of the background field. We have actually checked that using a  $512^3$   $N$ -body sample with nearly the same cosmology [the simulation is described in Colombi et al. (2007)] does not change any measurements presented in § 2.

Out of *FullMock*, we have extracted three spherical cuts of radius  $40 h^{-1}\text{Mpc}$  (hereafter denoted by *4k-mockX*), where the velocity-velocity comparisons are conducted, and twice deeper counterparts (hereafter denoted by *8k-mockX*) are used to give better constraints (§ 5.2) on the reconstruc-

tion within the volume of analyses. Each of these catalogues is centered in a different place in the simulation such that:

- *4k-mock6* is mildly overdense, with an effective mean matter density  $\Omega_{\text{eff}} = 0.35$ , and contains 495 haloes. It is designed in such a way that large voids and large concentrations of matter (clusters or super-clusters) are present near its boundaries, similarly as found in real redshift catalogues of our local neighbourhood, such as the UZC (Falco et al. 1999), the NBG-3k (Shaya et al. 1995; Tully et al. 2007) and the NBG-8k. This catalogue and its deeper counterpart, *8k-mock6*, are particularly suited to address edge effects on the NBG-3k (which terminates at Hydra and Centaurus clusters) and the NBG-8k (which stops at the Great Wall), respectively.

- *4k-mock7* is highly overdense, with  $\Omega_{\text{eff}} = 0.50$ , and contains 656 haloes. Very little mass has come in and out of this volume: it behaves somewhat like an isolated universe, with small external tides.

- *4k-mock12* is underdense, with  $\Omega_{\text{eff}} = 0.19$ , and contains 213 haloes. It presents as well a low level of density fluctuations along its boundary.

While there is no ambiguity in setting up a  $128^3$  MAK mesh when using all the haloes and the background particles (such as in *FullMock*), it is less trivial to consider lower resolution meshes that will be used in some of the subsequent analyses. Indeed, the number of mesh elements assigned to each tracer is not necessarily an integer anymore. Appendix A details the general procedure used to associate elements of the MAK mesh to each tracer.

## 2 ERRORS IN MAK VELOCITIES

Before going over observational issues, we address errors intrinsic to MAK reconstruction. First, there is scatter in the reconstruction of the displacement field itself which is expected to be rather small (Mohayaee et al. 2006). Second, there is scatter due to the Zel’dovich approximation one uses to convert a displacement field into a velocity field and to deal with redshift distortions. An accurate knowledge of the distribution of errors on the reconstructed velocities is eventually required for the likelihood analysis we want to introduce in § 6.2. In this section, we measure such a distribution in real space while redshift space will be addressed in § 4. In principle, the width of such a distribution is expected to increase when observational biases are taken into account while its shape should not change significantly.

We consider, in this section, reconstructions based on the catalog *FullMock*, for which periodic boundary conditions are applied to avoid edge effect problems. We also assume that we know the mass of all of described catalog objects (haloes and individual particles). Our subsequent reconstructions have a resolution within  $64^3$  and  $128^3$  mesh elements. We will thus present two reconstructions obtained on two different initial MAK mesh,  $128^3$  and  $64^3$ , obtained using the procedure presented in Appendix A. The results on the reconstructed displacement field are given in Fig. 3. These plots give the distribution of differences,  $P_{\text{DE}}$ , between the line of sight component of the reconstructed displacement field and the “exact” one, given by the simulation.

The dot-dashed and dashed curves correspond to a

least-square fit of the function  $P_{\text{DE}}$  corresponding to the  $128^3$  reconstruction respectively with a Gaussian fit, and a Lorentzian fit given by

$$P_{\text{Lor}}(x) = \frac{1}{\pi B} \frac{1}{1 + \frac{x^2}{B^2}}. \quad (4)$$

Examination of Fig. 3 supports the Lorentzian approximation with  $B = 35 \text{ km s}^{-1}$ , which reproduces better the long tails of  $P_{\text{DE}}$  than the Gaussian.

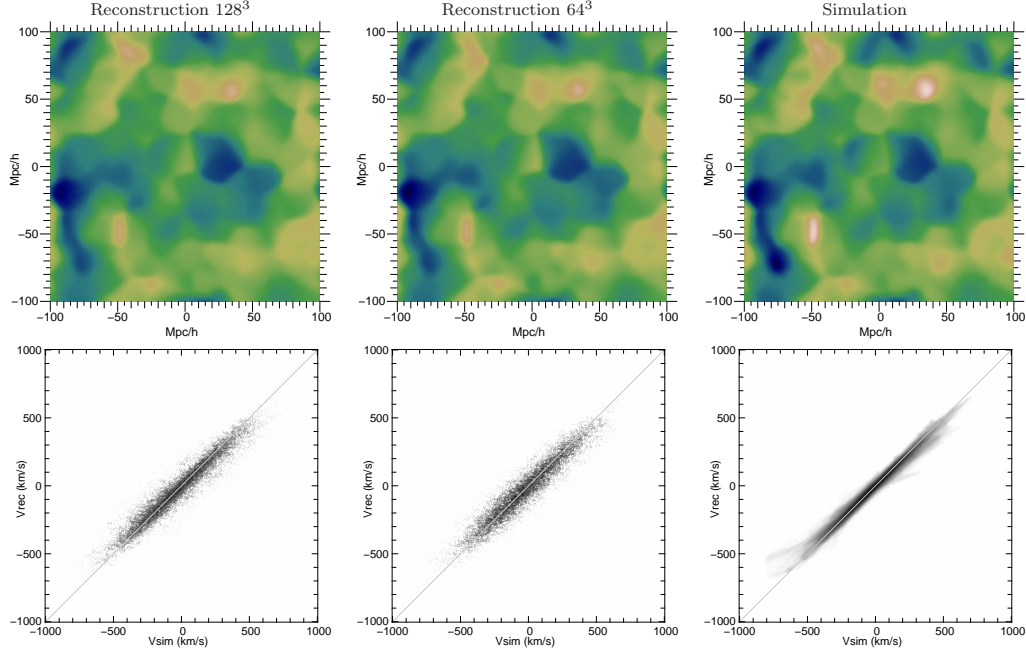
The width,  $B$ , of  $P_{\text{DE}}$  is rather small compared to the line-of-sight dispersion,  $\langle \beta^2 \Psi_r^2 \rangle^{1/2} \simeq 292 \text{ km s}^{-1}$ , as expected. Naturally, the function  $P_{\text{DE}}$  is slightly flatter and larger for the  $64^3$  case than for the  $128^3$  one. However, the far end tails of  $P_{\text{DE}}$  are the same for  $64^3$  and  $128^3$ . In this regime, the measurements are not influenced by the resolution of the grid used to perform the reconstruction but rather by the inability of MAK to reproduce the internal dynamics of massive, relaxed objects (Mohayaee et al. 2006).

Fig. 4 is similar to Fig. 3 but considers line of sight reconstructed velocities vs “exact” ones. Although Zel’dovich approximation introduces extra noise as shown by a wider width of the distribution,  $P_{\text{DE}}$  remains roughly Lorentzian with a small width  $B = 48 \text{ km s}^{-1}$ . This error variance is grossly 25% higher than the expected velocity field variance on the simulation volume (§ 1.1). We are thus not affected by cosmic variance effects that could have been induced by modes larger than the box size of the simulation.

These results are fully supported by the examination of Fig. 2. However, the lower panels of this figure shows that the joint distribution  $P(v_{\text{sim}}, v_{\text{rec}})$  presents non-trivial tails above the diagonal line in the lower left quadrant and below the diagonal line in the upper right quadrant, respectively. These tails do not disappear even after smoothing of the velocity field with a  $5 h^{-1} \text{ Mpc}$  Gaussian window. This is due to non-linear features in the dynamics not taken into account by our MAK+Zel’dovich prescription, which produces a slightly smoother velocity field than the real one. As a result, upper left panel of Fig. 2, which corresponds to the reconstruction, is less contrasted than the upper right one, which corresponds to the simulation.

These non-linear tails give a propeller shape to  $P(v_{\text{sim}}, v_{\text{rec}})$  which is susceptible to inducing a small bias on the final velocity-velocity comparison. For instance, one can estimate the slope of the lower left scatter plot of Fig. 2 using the ratio  $s = \sigma_{v,\text{rec}} / \sigma_{v,\text{sim}}$ , where  $\sigma_{v,\text{rec}}^2$  and  $\sigma_{v,\text{sim}}^2$  are the variances of the reconstructed and simulated velocity fields, respectively. In this case, the estimated  $\beta$  is biased to higher values by about 7%. However, visually inspecting the scatter shows no measurement bias should occur if only the central part of the scatter is used for the computation. To achieve this, we have first applied an adaptive SPH filter on the scatter plot to produce a Probability Density Function (PDF), which is probed by the scatter in the points, on a regular mesh grid. We then compute the  $1.5\sigma$  isocontour which encloses the region where the integrated PDF is equal to 68%. This procedure has already been used in Colombi et al. (2007) for the gravity-velocity comparison with total success. Only the points enclosed by the  $1.5\sigma$  isocontour are used to compute the new  $s_{\text{med},68}$  coefficient. The  $\beta$  parameter deduced from  $s_{\text{med},68}$  is now statistically unbiased. Similarly, we define two other slope estimators  $s_{\text{min},68}$  and  $s_{\text{max},68}$  whose relevance is discussed in Appendix B. In this





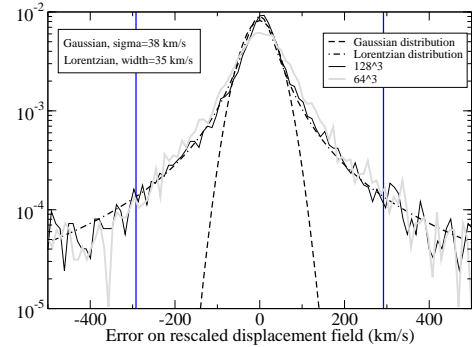
**Figure 2.** Velocity field reconstruction on *FullMock* – Top panels: A slice of the line-of-sight component of the simulated velocity field,  $v_{r,\text{sim}}$ , and the reconstructed one,  $v_{r,\text{rec}}$ , after smoothing with a  $5 h^{-1}\text{Mpc}$  Gaussian window. The observer is at the center of this slice. Bottom panels: Scatter plots between  $v_{r,\text{sim}}$  and  $v_{r,\text{rec}}$  for invidual haloes (left) and after smoothing (right).

paper, until § 6, we will only discuss the measurement of  $\Omega_m$  obtained through the estimation of  $s_{\text{med},68}$ . The  $\Omega_m$  obtained by this method is identified by a “1.5 $\sigma$ ” to make a difference with the one obtained through the likelihood analysis that will be established in § 6 and which is identified by a “ $\mathcal{L}$ ” in the tables and figures. A test of this method on a simulated scatter distribution, whose shape is built on analysis of reconstruction errors, is detailed in Appendix D.

### 3 MASS-TO-LIGHT ASSIGNMENT

Most reconstruction methods, including ours, infer the total matter distribution as a function of the visible matter distribution traced by galaxies. The fundamental assumption one usually makes is that the relation between these two distributions is highly deterministic. In other words, one assigns to each galaxy of a given luminosity  $L$  a dark matter concentration (a halo) of mass  $M = f(L)$ . However, there are several issues in this procedure:

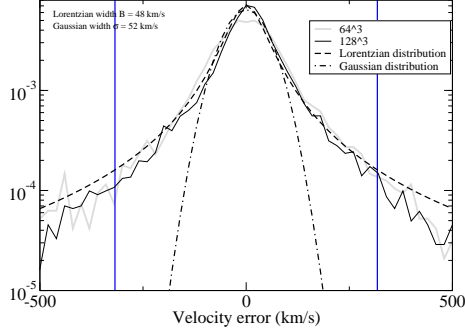
- *Mass-to-light ratio* – The choice of a function  $f(L)$  influences considerably the results and is expected to introduce significant bias on the measured  $\beta$  if performed unwisely. Now, the function  $f(L)$  is coarsely determined (Tully 2005; Marinoni & Hudson 2002) from direct measurements in observations. One way to infer this function is to rely on semi-analytic models of galaxy formation, but this represents a very strong prior on the measurements. Furthermore,  $f(L)$  remains a mean relation around which there can be some sig-



**Figure 3.** Error in reconstructed displacements – This plot displays the probability distribution of the quantity  $\beta(\Psi_{r,\text{rec}} - \Psi_{r,\text{sim}})$  measured in *FullMock* (solid curve), where  $\Psi_{r,\text{rec}}$  and  $\Psi_{r,\text{sim}}$  are the line-of-sight component of the reconstructed and simulated displacement fields, respectively, after choosing an observer at the center of the simulation box. The dashed and dot-dashed curves give the best fit of a Gaussian and a Lorentzian distribution, respectively.

nificant scatter. This dispersion can as well introduce some significant biases.

- *Missing tracers / Magnitude limitation* – Even if function  $f(L)$  is perfectly known, fainter galaxies are still miss-



**Figure 4.** *Error in reconstructed velocities* – Same as in Fig. 3 but the solid curve corresponds to the probability distribution of the quantity  $v_{r,rec} - v_{r,sim}$ , where  $v_{r,rec}$  and  $v_{r,sim}$  are the line-of-sight reconstructed and simulated velocities, respectively.

ing in the catalogues due to the limitations of observational instruments. For instance, in magnitude-limited catalogues, the number density of detected galaxies decreases with distance from the observer. These missing tracers have unknown positions and correspond to a part of the dark matter distribution which is totally undefined. This missing mass has to be taken into account in some way.

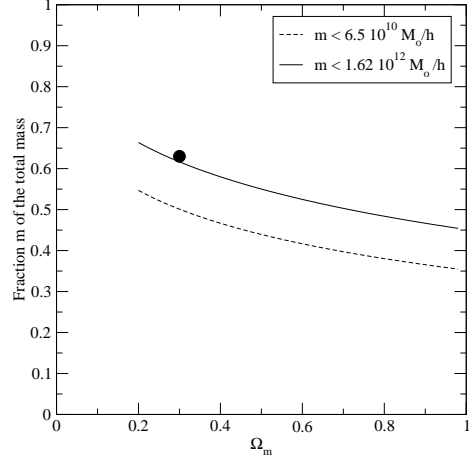
In what follows, we will first address the second issue in a very simple way which assumes that the function  $f(L)$  is well known (namely the masses of dark matter haloes themselves) but there is a fixed low-mass cut-off. The problem then consists in determining the unknown part of the dark matter distribution (namely the particles unassigned to any halo). Clearly it is correlated with the detected mass tracers but less clustered. There are two extreme ways to locate this missing mass

- (a) associate it with the existing tracers as usually done with the analysis of real observations
- (b) associate it with a uniform background.

Of course, the real solution is somewhat intermediate between (a) and (b) as will be shown in § 3.1.

Then, we turn in § 3.2 to the issue of the choice of  $f(L)$ . In this paper, we prefer to be as free as possible from strong priors so we deliberately do not use results from semi-analytic models of galaxy formation. Instead, we use determinations of  $f(L)$  from observational data but, unfortunately, there are large uncertainties in these measurements. The point here is to quantify, quite heuristically though, the effect of these uncertainties, random or systematic, on the measurement of  $\beta$ . Indeed, one is both confronted with a possibility of a wrong approximation of  $f(L)$  and most probably a large scatter around this mean relation.

In sufficiently deep galaxy catalogues, the effect of the missing tracers is expected to be negligible close to the observer and, in general, to increase with the distance from the observer. With appropriate weighting of the data, one can minimize the bias brought by the procedure used to infer the missing mass distribution far from the observer. In § 3.3, we shall illustrate this point by considering the case of



**Figure 6.** *Diffuse mass* – In this plot, we represent the fraction of the clustered mass below two mass resolutions for a standard  $\Lambda$ CDM type cosmology ( $h = 0.65$ ,  $\sigma_8 = 0.99$ ). We used a power spectrum as proposed by Bardeen et al. (1986). The curvature of the Universe is kept flat while  $\Omega_m$  varies. This fraction is plotted for mass resolutions:  $2.5 \times 10^{12} M_\odot$  (corresponding to the lower mass limit of haloes in our simulation) and  $10^{11} M_\odot$  ( $\simeq 10^9 L_{\odot,B}$ ). The unclustered fraction in *FullMock* is given by the back filled circle. The fraction of mass below both of these limits is still considerable.

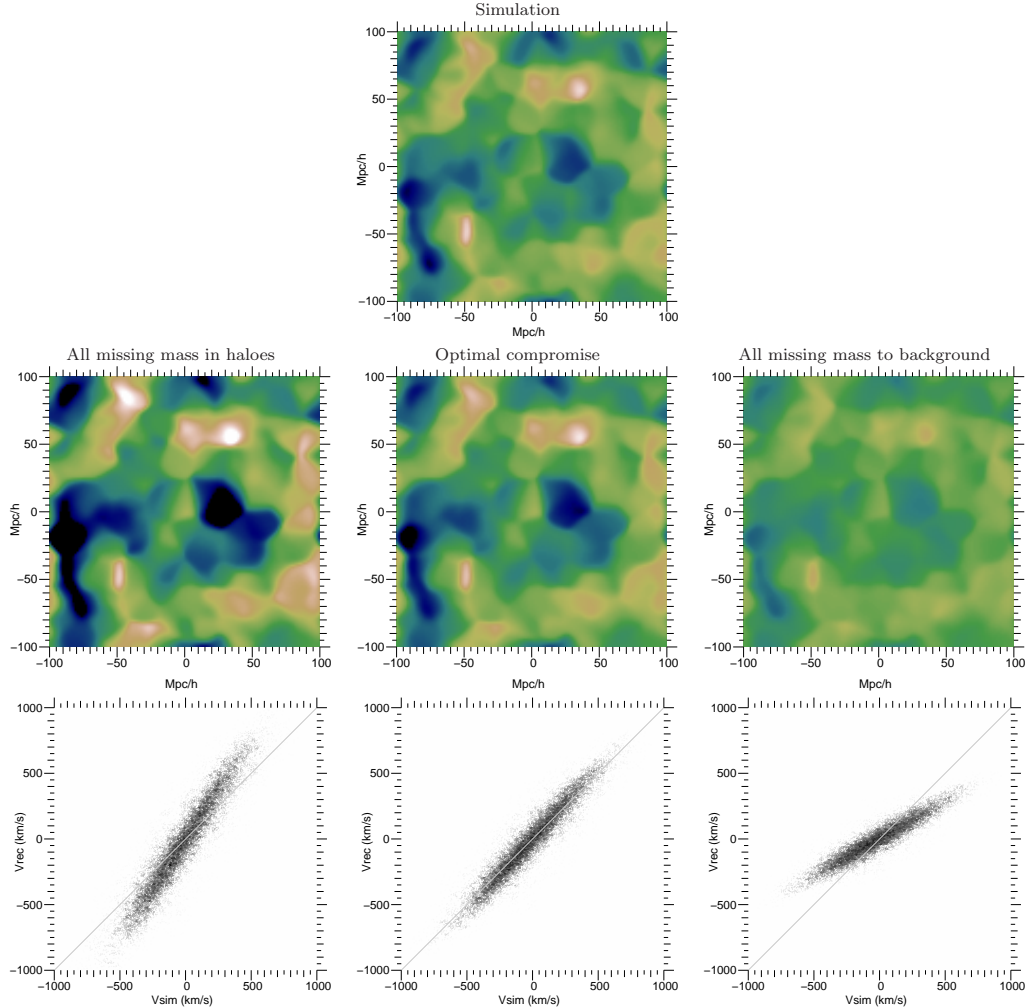
a magnitude-limited catalogue where all the missing mass is associated with the existing tracers [method (a) above].

### 3.1 Missing tracers

Fig. 6 shows the expected fraction of the total mass below a fixed threshold as a function of  $\Omega_m$ , using the Sheth & Tormen (2002) model (see also Fig. 1). The solid line corresponds to the mass cut-off of haloes in *FullMock* and agrees, as expected, with the measurement in the simulation for  $\Omega_m = 0.30$ . Here, 63% of the mass is outside of the haloes, which represent our “galaxies” with known  $M/L$  ratio. The particles not linked to the haloes represent the missing mass. In Fig. 2, their exact location was used to perform the reconstruction. The only information available now is the distribution of “visible galaxies”. The missing mass needs to be redistributed using only these pieces of information. We propose two extreme ways to do so:

I. *All missing mass to background* – Prior to the reconstruction, the missing mass is divided into particles which are randomly put in the catalog following a poissonian distribution. In the example illustrated by the right panels of Fig. 5 we choose for simplicity particles of the same mass as those in the simulation.

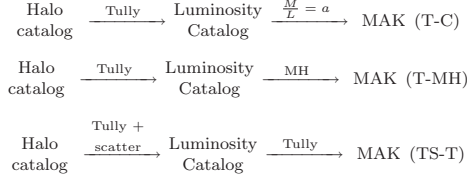
II. *All missing mass in haloes* – The missing mass is attributed to the existing haloes in proportion to their masses, as illustrated by left panels of Fig. 5. This approach is equivalent, in real observations, to multiplying the  $M/L$  ratio of galaxies or group of galaxies by a constant  $\alpha > 1$ .



**Figure 5.** *Diffuse mass correction* – The top panel gives a slice of the line-of-sight component of the simulated velocity field, after smoothing with a  $5 h^{-1}\text{Mpc}$  Gaussian window. The observer has been put at the center of this slice. The second row of panels represents the line-of-sight component of the reconstructed velocity field, smoothed in the same way, for different corrections of the diffuse mass. The third row of panels give the scatter distribution of individual reconstructed velocities of haloes *vs* simulated ones. The left panels give the result of a reconstruction on a mock catalogue which only contain the haloes and not the background field but at the same time conserves the total mass of the catalog by reassigning the missing mass to the haloes. The right panels give the result for a reconstruction based on a mock catalogue for which the missing diffuse mass is represented by a background field composed of particles placed randomly in the catalogue. The center panels give the result of a reconstruction on a mock catalogue which only contain the haloes and a random background field. The mass that have been initially removed from the mock catalogue (the background “galaxies”) is reassigned as follows: 60% to haloes and 40% to the background.

Obviously, in I, the screening effect due to the background is exaggerated, hence the reconstructed velocity is less contrasted and  $\beta$  is over-estimated to compensate for this. In II, on the other hand, the potential wells are more contrasted than they should be, which leads to the opposite effect. At this point, it is extremely tempting to try to find a

simple compromise between I and II as illustrated by middle panels of Fig. 5 where 60% of the missing mass was linked to the tracers and the remaining to a uniform background. With this particular choice of the redistribution, the match between the reconstructed and the simulated velocity fields is spectacular. This result is non-trivial given the simplic-



**Figure 7.** *M/L assignment* – Sketch of the procedures used to test the influence of a choice of a *M/L* assignment, as explained in the main text.

ity of the handling of this *sixty three* percent missing mass all the more since the scatter on the middle-lower panel of Fig. 5 is of the same order of that of the lower left panel of Fig. 2, where all the tracers contribute optimally.

Although the choice of the optimal redistribution remains a priori unknown in a real galaxy catalogue one can at least infer error bars from I and II. In that framework, Fig. 5 unfortunately provides quite a bad constraint on  $\beta$ ,  $0.36 \lesssim \beta \lesssim 0.85$ . However, in real galaxy catalogues, such as the NBG-3k or the NBG-8k, the minimum luminosity is of the order of  $10^9 L_\odot$ . This corresponds to a less abrupt mass cut-off,  $M_{\text{cut}} \sim 10^{11} M_\odot$ , than in Fig. 5, where  $M_{\text{cut}} = 2.5 \times 10^{12} M_\odot$ . Therefore, one expects the problem of missing mass to be less salient in real observations, as illustrated by the dashed curve of Fig. 6. Furthermore, an appropriate use of mock catalogues can help at calibrating the redistribution of mass, as performed in middle panels of Fig. 5.

### 3.2 Mass-to-light ratio

To test how the choice of mass assignment to galaxies or group of galaxies affects the results we consider the three following cases, as summarized in Fig. 7:

(i) *T-C case*: a galaxy catalogue is extracted from *FullMock* by associating a luminosity  $L(M)$  to each dark matter halo or background particle using Tully’s latest best fit of the group mass-luminosity relation (Tully 2005, see Fig. 8)

$$\frac{L_B}{L_\odot} = 2700 \left( \frac{M}{M_\odot} \right)^{0.59} e^{-6 \times 10^{11} M_\odot / M}, \quad (5)$$

which gives the luminosity in the B band for groups in the mass range  $10^{11} M_\odot - 10^{15} M_\odot$ . Then a new mass is given to each tracer assuming

$$M/L = \text{constant}, \quad (6)$$

as often used in the literature, and MAK reconstruction is performed on a resampling of this mass distribution.

(ii) *T-MH case*: a less extreme case than assuming  $M/L = \text{constant}$  consists in separating the tracers in three broad classes: faint galaxies, luminous galaxies and group/clusters of galaxies, as performed by Marinoni & Hudson (2002), hereafter MH. To do this, they used a simple mapping between the Schechter luminosity function and the

Press-Schechter mass function that reads as follows

$$\begin{aligned} M/L &= 1.15 \cdot 10^7 \left( \frac{L}{L_\odot} \right)^{-0.5} h \frac{M_\odot}{L_\odot} & \frac{L}{L_\odot} < 4 \cdot 10^{10} \\ M/L &= 128 h \frac{M_\odot}{L_\odot} & 4 \cdot 10^{10} < \frac{L}{L_\odot} < 4 \cdot 10^{11} \\ M/L &= 3.6 \cdot 10^{-4} \left( \frac{L}{L_\odot} \right)^{0.5} h \frac{M_\odot}{L_\odot} & \frac{L}{L_\odot} > 4 \cdot 10^{11} \end{aligned} \quad (7)$$

as shown in upper panel of Fig. 8. In this framework, we generated the same catalog as in *T-C* case but it was analyzed assuming the *M/L* function given by Eq. (7).

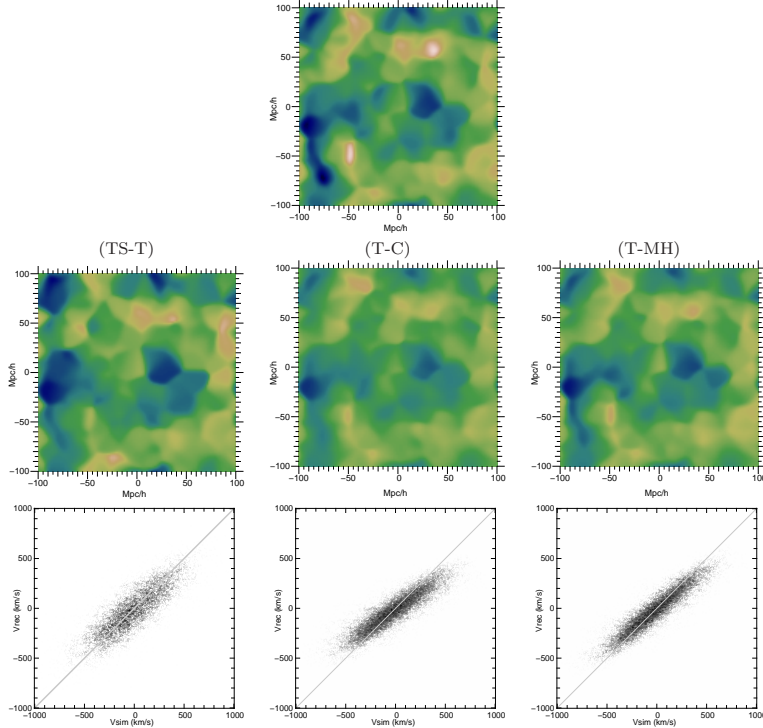
(iii) *TS-T case*: assuming that we have an unbiased estimator of the *M/L* function, there can still be a scatter around this mean value that can increase the errors and also introduce systematic bias. We test this by multiplying the mass of each halo of *FullMock* by a random number  $x$  such that  $\log_{10} x$  is uniformly distributed in  $[-1, 1]$ , prior to MAK reconstruction, which is performed on a resampling of the halo catalog following the procedure explained in Appendix A. Note that the mass of background particles remains unchanged during the process, which corresponds to 63% of the matter distribution being unaffected by the scattering. However, applying the scatter to small mass haloes only introduces a local additional noise which should not have any significant consequences on the reconstruction accuracy for which deeper potential wells are in fact more critical.

We want to highlight the fact that each of these transformations, actually corresponding to transforming the mass of an object of *FullMock* through a  $M \rightarrow L \rightarrow M$  operation, does not correspond to an identity. One actually gets a new set of masses attached to each tracer which is different from the original one. Moreover, the output mass distribution  $P_{\text{mass,out}}(M)$  may be fundamentally different from the input one  $P_{\text{mass,in}}(M)$ . Indeed, computing  $P_{\text{mass,out}}(M)$  is equivalent to performing a weighted average of  $P_{\text{mass,in}}(M)$ . This procedure induces a global reshaping of the distribution. Consequently, the statistical properties of the corresponding mass density field may be affected.

More technically, during the procedure used to construct all the catalogues above, total mass conservation is enforced. Note that the total mass depends on  $\Omega_m h^2$ , but this normalization does not affect MAK displacements, which are sensitive to density contrasts only. Parameters  $\Omega_m$  and  $h$  in fact intervene while performing velocity-velocity comparison and while converting distances to velocities (§ 6), respectively.

As expected, random uncertainty on the mass determination does not introduce any bias, it only increases the scatter in the measurements as illustrated by the lower left panel of Fig. 9. A more important issue is the global knowledge of the *M/L* relation. Indeed, it seems that the slope of this relation influences greatly the results, as illustrated by the middle and right panels of Fig. 9. Clearly, if the galaxies follow the Tully formula (5), it is definitely wrong to assume constant *M/L* and even the MH fit introduces a significant bias, although it is well within the observational errors compared to Eq. (5). It must be noted that this bias can be turned into an advantage if one does not want to measure  $\Omega_m$  but the *M/L* relation. Indeed, WMAP experiment (Bennett et al. 2003; Spergel et al. 2006) coupled with

## Observational biases in Lagrangian reconstructions of cosmic velocity fields 9



**Figure 9.** *M/L bias* – The top panel gives the expected line-of-sight component  $v_r$  of the velocity field, smoothed with a  $5 h^{-1}\text{Mpc}$  Gaussian filter, as given by the simulation in a thin slice of the simulation containing the observer. The middle panels gives the reconstructed  $v_r$  field, with the same smoothing, after having applied each of the transformations specified in Fig. 7 to *FullMock*. The lower panels gives the scatter between the reconstructed and simulated peculiar velocities for each of the transformations.

**Table 1.** *M/L bias effect* – This table gives the results obtained using different statistical tools. We also measured  $\Omega_m$  using six different methods: the label  $s$  means we used the slope estimated by using all objects, the label  $\mathfrak{L}$  is used when  $\Omega_m$  has been determined using the likelihood analysis, and the label  $1.5\sigma$  is used when the slope is estimated using only the objects within the  $1.5\sigma$  isocontour of the PDF between reconstructed velocities and simulated velocities (method described in § 2).

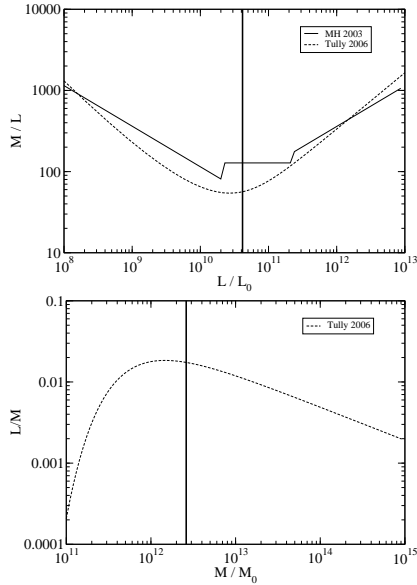
Transf.	Velocity			$\Omega_m$ ( $s$ )	$\Omega_M$ ( $\mathfrak{L}_{\min}$ )	$\Omega_M$ ( $\mathfrak{L}_{\max}$ )	$\Omega_M$ ( $1.5\sigma, s_{\text{med}}$ )	$\Omega_M$ ( $1.5\sigma, s_{\min}$ )	$\Omega_M$ ( $1.5\sigma, s_{\max}$ )
	$s$	$r$	$\sigma$						
None	0.88	0.89	0.58	0.38	0.30	0.31	0.30	0.28	0.31
TS-T	0.90	0.78	0.64	0.36	0.26	0.30	0.28	0.24	0.33
T-MH	0.80	0.80	0.60	0.45	0.33	0.38	0.36	0.32	0.40
T-C	0.71	0.78	0.63	0.55	0.40	0.48	0.44	0.37	0.51

an analysis of the power spectrum of large scale density of galaxies (Tegmark et al. 2006) gives good constraints on the real  $\Omega_m$  now. Our method, on the other hand, is able to measure the discrepancy between the measured  $\beta$  and the expected growth factor  $\beta_{\text{expected}} = \Omega_m^{5/9}$  (*i.e.* the bias). This measurement may give an idea of how wrong is the assumed  $M/L$  relation prior to the reconstruction and may push us to try different plausible  $M/L$  functions. Thus our method is able to measure the *way* that the matter is distributed in the Universe once it is given its average density  $\Omega_m$ . On the other hand, if the above bias is well understood, this method

helps at reducing the degeneracy in the determination of cosmological parameters. Indeed, our posterior probability on  $(\Omega_m, h)$  gives a constraint orthogonal (for example see the results in Mohayaee & Tully 2005) to the one obtained from the WMAP experiment and from the galaxy statistics of the SDSS.

### 3.3 Magnitude limitation

Magnitude-limited sampling of mass tracers introduces a new type of problem: flux limitation decreases the mass res-



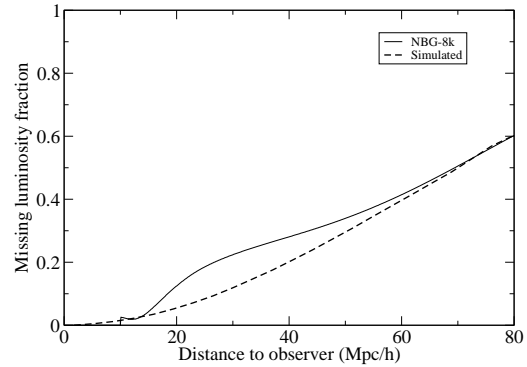
**Figure 8.** *M/L function* – The above two plots give the forward and inverse mass-to-light functions for both Tully (2005) and Marinoni & Hudson (2002) fits. The top panel gives the  $M/L$  as a function of the luminosity  $L$ , the bottom panel gives  $L/M$  as a function of the mass  $M$ .

olution toward the outer edges of the catalogue contrary to the homogeneous case studied in § 3.1. Usually, the incompleteness is handled by boosting uniformly the luminosities of galaxies at a given distance from the observer (Branchini et al. 2002), prior to conversion of luminosities into masses. This is a fair approach if  $M/L = \text{constant}$ , modulo the issues discussed in § 3.1. However, this method is in general questionable for non-trivial  $M/L$  relations as in Eq. (5) or if different  $M/L$ 's are assigned to galaxies with different types. In these last two cases, the missing mass correction should be applied to the mass distribution itself instead of the luminosity one, to avoid systematic errors on mass assignments, hence on reconstructed velocities. This unfortunately requires a prior assumption on the value of  $\Omega_m$ , but only slightly complicates the analyses.

In the observational data, galaxies are separated into two populations: groups<sup>4</sup> of galaxies (Tully 1987) and field galaxies. These two populations should be treated separately, keeping in mind that the groups are the most critical because their gravitational influence is much larger than individual field galaxies and they have better peculiar velocity measurements.

The full procedure consisting of creating a magnitude-limited mock catalogue and recovering the mass distribution is detailed in Appendix C. Let us recall that, in our mock catalogues, groups of galaxies are simulated dark matter haloes with more than 5 particles while background galax-

<sup>4</sup> Groups are defined here as compact sets of 5 galaxies or more.



**Figure 10.** *Magnitude limitation* – Solid line: NBG-8k predicted luminosity incompleteness at the given distance from the observer. Dashed line: Simulated luminosity incompleteness in *8k-mock6*. The incompleteness is expressed in terms of missing luminosity fraction at the specified distance.

ies are identified with dark matter particles unassigned to any halo. We list here the key steps we used to correct for incompleteness:

- I. The total apparent luminosities of groups of galaxies is obtained assuming a global or a local Schechter luminosity distribution for the considered groups. The intrinsic luminosity is computed trivially from the total apparent luminosity and the redshift of the group.
- II. The intrinsic luminosity of the remaining unbound galaxies (thus field galaxies) is also determined, straightforwardly.
- III. Then, masses are estimated by assigning appropriate  $M/L$  to each object of I and II.
- IV. The local missing mass from undetected background galaxies is inferred from the detected mass distribution. This requires a prior on  $\Omega_m$ .
- V. This missing mass may either be reassigned locally to detected field galaxies of II (our choice) or be introduced by the mean of new randomly positioned tracers, as discussed in § 3.1.

To examine the effects of systematics in the correction for incompleteness, we use *8k-mock6* and choose a flux limit such that the resulting mock catalogue has an incompleteness similar to NBG-8k, as shown in Fig. 10. Results are summarized in Fig. 11 and in Table 2.

The reconstructed radial peculiar velocities  $v_{r,\text{rec}}$  are behaving extremely well. On average, the comparison between simulated and reconstructed velocity fields is surprisingly good in a volume of radius  $80 h^{-1}\text{Mpc}$ , even though the edge misses locally 98% of the field galaxies which represents 60% of the total mass in our mock catalogue. It means that, though we keep only 2% of the field galaxies, they suffice, in addition to the groups, for a reasonably fair recovery of the large-scale peculiar velocity field. Note the small bias in the scatter of the lower right panel of Fig. 11, resulting in a slightly larger  $\Omega_m = 0.38$  than the expected value of 0.30, but in good agreement with the effective value of 0.35 expected in the corresponding volume (see § 5.3 on cosmic



variance effects). This bias might be the consequence of our treatment of the missing mass coming from undetected tracers as discussed, in detail in Appendix C (point B).

#### 4 REDSHIFT DISTORTION

The input of MAK reconstruction is the position of objects in real space as needed by Eq. (2). However redshift catalogues give us galaxy positions in redshift space, namely  $s_r = Hd + v_r$ , where  $s_r$  is the redshift distance,  $d$  is the luminosity distance between the observer and the object and  $v_r$  is the line-of-sight peculiar velocity. To account for redshift distortions, we must correct for two major effects:

- “Fingers-of-god” correspond to an elongation of dense structures along the line of sight, such as clusters of galaxies, due to random motions of galaxies within these structures.
- Kaiser effect (Kaiser 1987) is a large-scale effect coming from the coherent part of the cosmic flows, which, for instance, increase the overall density contrast.

Fingers-of-god effects can be easily removed by simply collapsing groups or clusters to a single point, as usually performed in the literature. However, such a procedure is generally carried out in a rather ad-hoc way and is certainly not free of biases.

The Kaiser effect can be accounted for by modifying the cost function (2) using the Zel’dovich approximation to infer line-of-sight peculiar velocities as functions of the sought displacement field (Mohayaee & Tully 2005; Valentine et al. 2000). If  $\mathbf{s}(\mathbf{q})$  is the redshift coordinate of a particle originally at  $\mathbf{q}$  then the total cost (2) of the association  $\sigma$  becomes:

$$I_\sigma = \sum_{i=1}^N \left( (\mathbf{s}_i - \mathbf{q}_{\sigma(i)})^2 - \frac{\beta(2+\beta)}{(1+\beta)^2} \frac{((\mathbf{s}_i - \mathbf{q}_{\sigma(i)}) \cdot \mathbf{s})^2}{\|\mathbf{s}\|^2} \right), \quad (8)$$

where  $\beta$  is the linear growth factor. Once the redshift displacement  $\Psi^s = \mathbf{s} - \mathbf{q}$  has been computed, the reconstructed radial peculiar velocity of the object  $i$  can be obtained by

$$v_{r,\text{rec}}^s = \frac{\beta}{1+\beta} \frac{\mathbf{s} \cdot \Psi^s}{\|\mathbf{s}\|}. \quad (9)$$

The cost function  $I_\sigma$  leads to the exact result in the case of a Zel’dovich displacement field without shell crossing after redshift distortion. However, in general, the second term (accounting for redshift distortion) of Eq. (8) becomes of the same order as the first term (the real space cost term) near the origin. In this case, the reconstruction becomes ill-defined because of the loss of convexity of functional  $I_\sigma$ . We expect thus the central part of all catalogues to be, in general, poorly reconstructed. The size of such a region is roughly determined by the magnitude  $v_{\text{obs}}$  of the large-scale flow nearby the observer with respect to the Cosmic Microwave Background. The velocity  $v_{\text{obs}}$  determines the relative contribution of the first term with respect to the second term of Eq. (8). In practice  $v_{\text{obs}}$  is of the order of a few hundred  $\text{km s}^{-1}$  (for instance the Local Group velocity is

630  $\text{km s}^{-1}$ , Erdoğan et al. 2006) which gives us a region of “exclusion” of radius of about a few  $h^{-1}\text{Mpc}$ .<sup>5</sup>

Again, MAK reconstruction fails in regions where shell crossings occur. Projection in redshift space generates such shell crossings along the line-of-sight. These shell crossings are dramatic because of their anisotropic nature. In particular, filaments can cross each other while passing from real to redshift space, implying the reconstruction will fail in a large region of the catalogue encompassing the gravitational influence of these filaments. In this area, most of the reconstructed radial velocities will have the opposite sign compared to the true velocity. Of course, shell crossings in redshift space can have more complex consequences but this simple example suggests that MAK reconstruction should not work as well in redshift space as in real space.<sup>6</sup>

Another problem of this method is that one must assume  $\beta$  prior to the reconstruction. As for § 3.3, where we had to guess the undetected mass, we choose a value  $\Omega_{m,\text{in}}$ , thus an assumed  $\beta_{\text{in}}$ , then we make a redshift reconstruction and measure a  $\Omega_{m,\text{out}}$ . In practice, the “true”  $\Omega_m$  of the catalogue was chosen to be the one for which  $\Omega_{m,\text{in}} = \Omega_{m,\text{out}}$ , which corresponds to having self-consistent orbits modeling when doing MAK reconstruction and when one makes a comparison with measured velocities.

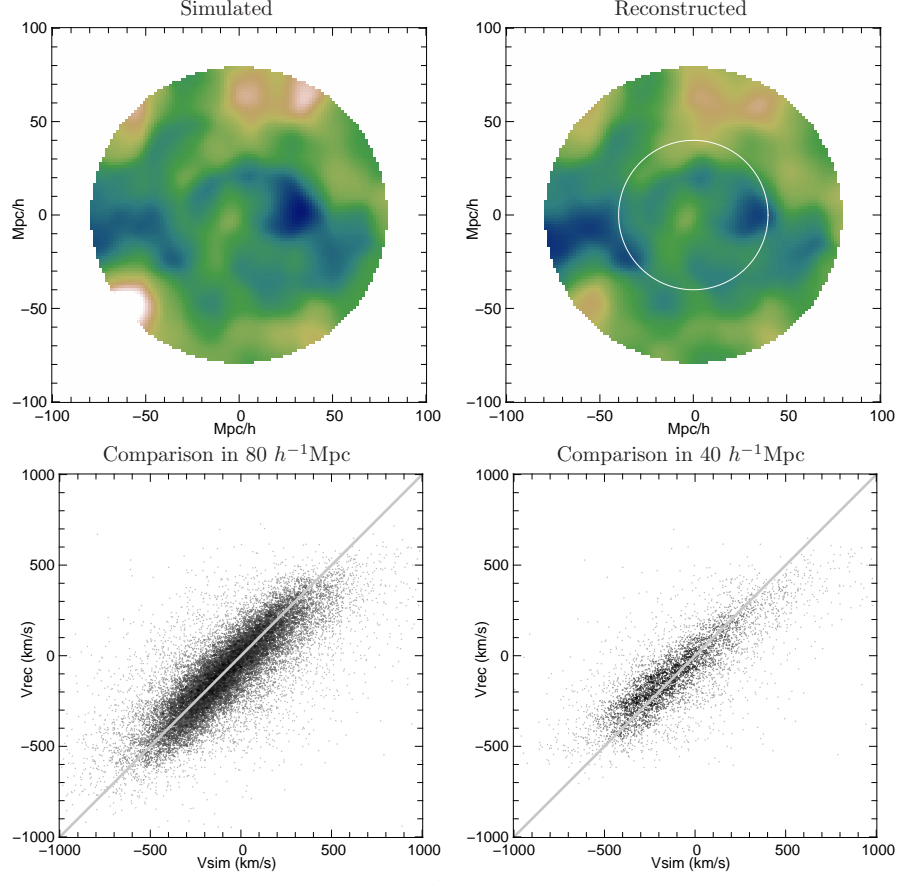
Fig. 12 shows both reconstructed and simulated velocity fields and the scatter between  $v_{r,\text{rec}}^s$  and  $v_{r,\text{sim}}$ . The first impression when comparing the two top panels of Fig. 12 is that the redshift reconstruction behaves really well. However, some potentially worrying localized features are present:

- Some important structures have their velocities badly reconstructed. Two important examples are the green-yellowish finger just above the center of the upper right panel of Fig. 12 and the big velocity peak at the top of this same panel. In the left panel, these two structures are not so prominent. The difference can be understood by studying the impact of the Kaiser effect on the reconstructed velocity field. Basically, two nearby filaments can merge in redshift space and give birth to a filament with a higher apparent density. The reconstruction is not able to separate these two filaments, which leads to an area with higher reconstructed velocities than the true ones. Thus, we expect in observational data to meet problems in the neighbourhood of the Great Wall, which is a supercluster of filaments compressed by redshift distortion.

- The velocity field in the immediate (5-10  $h^{-1}\text{Mpc}$ ) neighbourhood of the mock observer has lost its spatial structure and even presents a spurious peak. This is, most unfortunately, an expected problem that is linked to the above discussion on the problems of  $I_\sigma$  near the observer. Indeed, in the neighbourhood of the observer,  $I_\sigma$  becomes singular and the reconstruction misses, most likely, the right orbits. Analysing the smoothed velocity field seems to show that this effect looks in practice much like the one just above: the reconstructed velocity field may be boosted by the merging of different structures in the neighbourhood of the observer.

<sup>5</sup> See e.g. Colombi et al. (2007) for a similar discussion.

<sup>6</sup> This is also true for the Least-Action method for which multiple solutions quickly arises.



**Figure 11.** *Incompleteness: magnitude limitation* – Top panels: A slice of the line-of-sight component of the simulated velocity field in 8k-mock6 and the reconstructed one, after smoothing with a  $5 h^{-1}\text{Mpc}$  Gaussian window. The displayed slice is chosen to include the observer in  $(0, 0)$ . The white circle in the right panel gives the size of the  $40 h^{-1}\text{Mpc}$  sphere embedded in the  $80 h^{-1}\text{Mpc}$  one. Bottom panels: The scatter plots compare the reconstructed and simulated velocities of objects in the  $80 h^{-1}\text{Mpc}$  region (left panel) and in the  $40 h^{-1}\text{Mpc}$  volume (right panel).

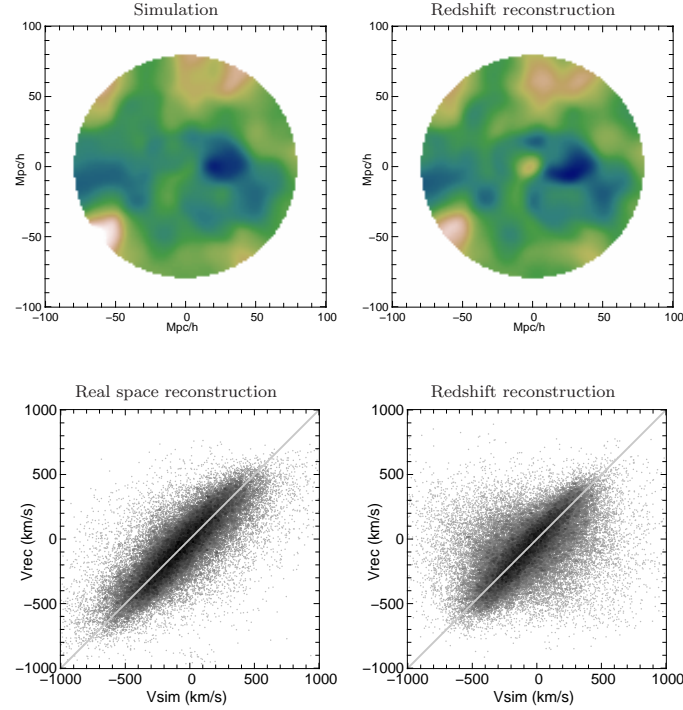
**Table 2.** *Incompleteness: magnitude limitation* – Column description is given in the caption of Table 1.

Volume	Velocity field			$\Omega_m$ (s)	$\Omega_m$ ( $\mathcal{L}_{\min}$ )	$\Omega_m$ ( $\mathcal{L}_{\max}$ )	$\Omega_m$ ( $1.5\sigma, s_{\text{med}}$ )	$\Omega_m$ ( $1.5\sigma, s_{\min}$ )	$\Omega_m$ ( $1.5\sigma, s_{\max}$ )
	s	r	$\sigma$						
8k	0.86	0.77	0.64	0.39	0.26	0.31	0.29	0.25	0.34
4k	0.77	0.75	0.66	0.48	0.37	0.45	0.38	0.30	0.47

**Table 3.** *Redshift reconstruction* – Column description is given in the caption of Table 1.

s	r	$\sigma$	$\Omega_m$ (s)	$\Omega_m$ ( $\mathcal{L}_{\min}$ )	$\Omega_m$ ( $\mathcal{L}_{\max}$ )	$\Omega_m$ ( $1.5\sigma, s_{\text{med}}$ )	$\Omega_m$ ( $1.5\sigma, s_{\min}$ )	$\Omega_m$ ( $1.5\sigma, s_{\max}$ )
0.83	0.46	0.95	0.50	0.22	0.29	0.27	0.22	0.33





**Figure 12.** *Redshift distortion correction* – Top panels: A slice of the smoothed velocity field  $v_{r,\text{sim}}$  and  $v_{r,\text{rec}}$  is shown in the left and right panels, respectively. The two fields have been smoothed with a  $5 h^{-1}\text{Mpc}$  Gaussian window, with objects put at their real (simulated and reconstructed) comoving coordinates. Bottom panels: Scatter plots between  $v_r$  and  $v_{r,\text{rec}}$  for individual mass tracers. The left panel (right panel) was produced using a real space reconstruction (redshift space reconstruction, respectively). In both cases, only objects within a sphere of  $8000 \text{ km s}^{-1}$  are shown.

- The lower right panel presents two additional off-diagonal tails compared to lower left panel. As discussed earlier, these tails are due to shell-crossings occurring along the lines-of-sight when passing from real to redshift space. These extra shell-crossings result in some reconstructed velocities acquiring a sign opposite to the true velocities.

Similarly as in § 2, we have computed in Fig. 13 the distribution of differences  $P_{\text{VE}}^s$  between  $v_{r,\text{rec}}^s$  and  $v_{r,\text{sim}}^s$ , for a redshift reconstruction applied on 8k-mock6 based on a  $64^3$  mesh.<sup>7</sup> Though the distribution is of course wider than in Fig. 4, the previously drawn conclusions are still valid.  $P_{\text{VE}}^s$  is better fitted by a Lorentzian distribution with  $B = 86 \text{ km s}^{-1}$  than by a Gaussian of width  $\sigma = 91 \text{ km s}^{-1}$ , particularly in the tails.

To check the effects of redshift distortion on the quality of the reconstruction, one can compare Table 3 to the first row of Table 1. As usual, the  $s$  parameter is slightly biased below unity due to nonlinear effects discussed in § 2, which seem, not surprisingly, to be slightly enhanced by redshift distortions. The appearance of the off-diagonal tails in the

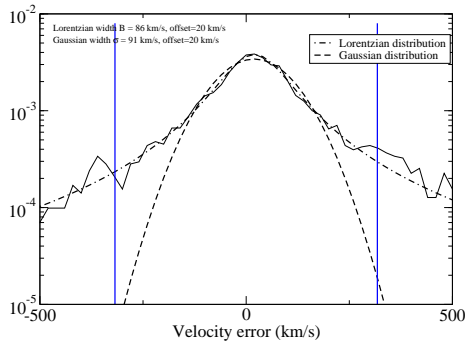
lower right panel of Fig. 12 increases the level of scattering, hence the correlation coefficient  $r$  decreases and the signal-to-noise ratio increases. Reducing the analysis to the region inside  $1.5\sigma$  isocontour greatly improves the results, as expected, but still leads to a value of  $\Omega_m$  slightly biased to lower values,  $\Omega_m = 0.27$ .

## 5 EFFECTS OF CATALOGUE GEOMETRY

In practice, real galaxy catalogues are not spatially periodic as is our simulation. They represent a region of finite volume with non-trivial geometry. In particular, two kinds of problems arise:

- *Edge effects* – Reconstruction of the galaxy trajectories without any piece of information on what may affect them dynamically from the outer parts of the catalogue is likely to introduce significant sources of errors, possibly systematic. We separate here edge effects into two subclasses: the effects of the obscuration by our galaxy, which defines a Zone of Avoidance (hereafter ZOA) and the effects of finite depth of the catalogue. These two effects need a separate treatment detailed in § 5.1 and § 5.2.

<sup>7</sup> The handling of the finiteness of the catalogue volume is handled in § 5.2



**Figure 13.** Error distribution of the reconstructed velocity field, redshift space – Same as in Fig. 4 but the solid curve corresponds to the probability distribution of the quantity  $v_{r,rec}^s - v_{r,sim}$ , where  $v_{r,rec}^s$  and  $v_{r,sim}$  are the line-of-sight reconstructed and simulated velocities, respectively.

- *Cosmic variance* – The finite volume of the accessible part of the Universe might be a potentially unfair realization of the random process underlying the properties of the large scale matter distribution. We must investigate whether our method, including handling of edge effects, is robust to the recovering of the statistical properties of the whole Universe from observations of only a fraction of it.

### 5.1 Zone of avoidance

Dust present in the Milky Way’s galactic plane highly attenuates the light, thus galaxy catalogues generally do not provide any data in this direction (approximately the region within  $|b| < 5$  deg, where  $b$  is the galactic latitude) of the ZOA. This strong attenuation introduces a boundary effect, which has the unpleasant feature of being present at any distance from the observer and may thus severely affect the measurements. As this area is nonetheless relatively small, particularly at low redshift, a simple correction should be able to greatly remove the boundary effect in the inner region of the catalogue.

Simulating the effect is made easy by putting an observer at the center of the simulation volume and by removing all mass tracers in the neighbourhood of the galactic plane  $z = 0$ , *i.e.* which have  $|b| < \alpha$ . This gives us *FullMock-ZOA*.<sup>8</sup>

Though more advanced ways of filling the ZOA exists (e.g., Lahav et al. 1994; Fontanot et al. 2003), this latter is here sufficiently small to be dealt with by the following simple algorithm. Since the statistical properties of the galaxies should not change across the boundaries of the ZOA, the objects in its neighbourhood can be used to fill the zone. We build new mass tracers to fill the obscured area by applying a locally planar symmetry transformation to the galaxies and groups with  $-3\alpha < b < -\alpha$  according to the “plane”  $-\alpha$ . We execute the same operation on objects with  $+\alpha < b < +3\alpha$  but according to the “plane”  $+\alpha$ . In the end, the masses of

the copied haloes in the ZOA are divided by two and we only take half of the field galaxies. This method has been used previously to fill the zone of avoidance in NBG-3k (Shaya et al. 1995) and NBG-8k. This folding procedure has been applied to *FullMockZOA*, slightly moving some of the newly created objects to enforce the periodicity of the simulation box to avoid mixing the effect of the ZOA with other boundary effects. The results are presented in Fig. 14. As expected, the ZOA has a clear impact on errors of the reconstructed velocities.

The typical errors on the reconstructed velocities, represented in the left panel of this figure, rise substantially in the vicinity of the obscured area. Fortunately, they remain well below the natural velocity dispersion of the simulation (dashed line). As we are comparing velocity fields filtered with a  $5 h^{-1}\text{Mpc}$  Gaussian window, we expect the reconstructed velocity field to be nearly error free for all points nearer than about  $60 h^{-1}\text{Mpc}$ .<sup>9</sup> It is also fortunate we have not introduced an extra bias using the filling algorithm, as shown both by comparing Table 4 to the first row of Table 1 and looking at the scatter plot in the right panel of the Fig. 14. We nonetheless highlight that the edge effect is not at all localized near the ZOA but extends quite far away and becomes negligible only for  $|b| > 20$  deg. Table 4 shows that the above extra noise does not have any impact on the measured  $\Omega_m$ .

### 5.2 Lagrangian domain

The inputs to MAK reconstruction are the present coordinates of the objects, *i.e.*  $\mathbf{x}$  in Eq. (2) or  $\mathbf{s}$  in Eq. (8), and the knowledge of the Lagrangian domain, *i.e.*  $\mathbf{q}$  in Eq. (2) or (8). Redshift catalogues give the present “positions” of the objects, *i.e.*  $\mathbf{s}$  in Eq. (8), however we have no observations that would give us the corresponding Lagrangian domain  $\mathbf{q}$ . We are thus limited to make guesses, though in the end, for huge catalogues, the details of the guess does not matter as gravitational forces are screened on large scales by the nearly homogeneous distribution of matter in the universe. Consequently, what happens at the boundaries should not strongly affect the central part of the catalogue though some guesses may be better at confining the edge effects on the boundaries. The naive solution is to assume that the Lagrangian domain is not so different from the volume of the catalogue itself. This assumption only begins to be a good approximation for volume enclosed in a sphere for which radius is big enough. For our  $80 h^{-1}\text{Mpc}$  sample, the mass going in and out of the volume (from initial to present time) already represents about 16% of the total mass. For a  $40 h^{-1}\text{Mpc}$  sphere, the mass flow is even greater: it may vary between 30% and 63 % of the total mass depending on the 8k-mock catalogue considered. Though tidal field and cosmic variance effects becomes negligible on a  $80 h^{-1}\text{Mpc}$  scale, they still affect the boundaries of the Lagrangian domain of a given catalogue in a non-trivial way. As we shall show these problems are further enhanced by redshift distortion.

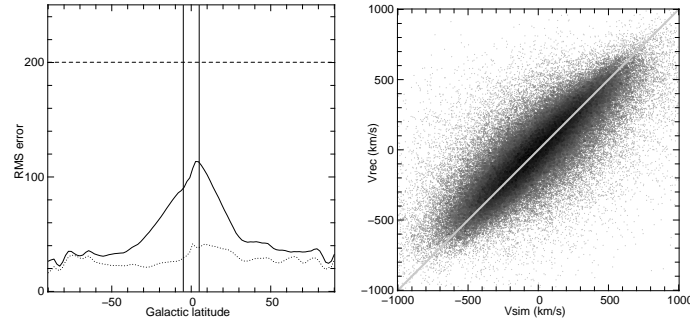
To achieve a meaningful comparison, we have run a reconstruction on 8k-mock6 using the Lagrangian domain

<sup>8</sup>  $\alpha = 5$  deg in our case.

<sup>9</sup> This corresponds to taking a 5 deg wide ZOA and computing at what distance the window is smaller than the ZOA.

**Table 4.** Zone of avoidance – Noise and biasing summary. Column description is given in the caption of Table 1.

$s$	$r$	$\sigma$	$\Omega_m(s)$	$\Omega_m(\mathfrak{L}_{\min})$	$\Omega_m(\mathfrak{L}_{\max})$	$\Omega_m(1.5\sigma, s_{\text{med}})$	$\Omega_m(1.5\sigma, s_{\min})$	$\Omega_m(1.5\sigma, s_{\max})$
0.89	0.79	0.61	0.37	0.30	0.35	0.32	0.285	0.36

**Figure 14.** Zone of Avoidance / velocity field – Left panel: Binned average RMS (Root Mean Square) error on the smoothed velocity field. As usual, the velocity field window has been smoothed with a 5 Mpc/h Gaussian window. Each point is computed by averaging the square of the deviation of the velocity field along the line-of-sight and for all line-of-sights having belonging to the same  $\sin(b)$  bin, where  $b$  is the “galactic” latitude. The solid line gives the RMS error in the presence of a zone-of-avoidance at  $b = 0$ . The dotted line gives the RMS error for a reconstruction on a catalogue without ZOA. The dashed line gives the RMS of the smoothed velocity field itself. Right panel: Scatter plots between  $v_{r,\text{rec}}$  and  $v_{r,\text{sim}}$  for individual mass tracers.

given by the simulation; this reconstruction is called *TrueDom*. Now, we confront the results of *TrueDom* for two different reconstruction setups that try to recover the Lagrangian domain:

- *NaiveDom* reconstruction is obtained by assuming a naive spherical Lagrangian domain for 8k-mock6. In that case, all the mass that is presently in the 8k-mock6 catalogue was uniformly in a sphere of radius  $80 h^{-1}\text{Mpc}$ . Equivalently, it means no significant mass flow must have gone through the comoving boundaries in the past.

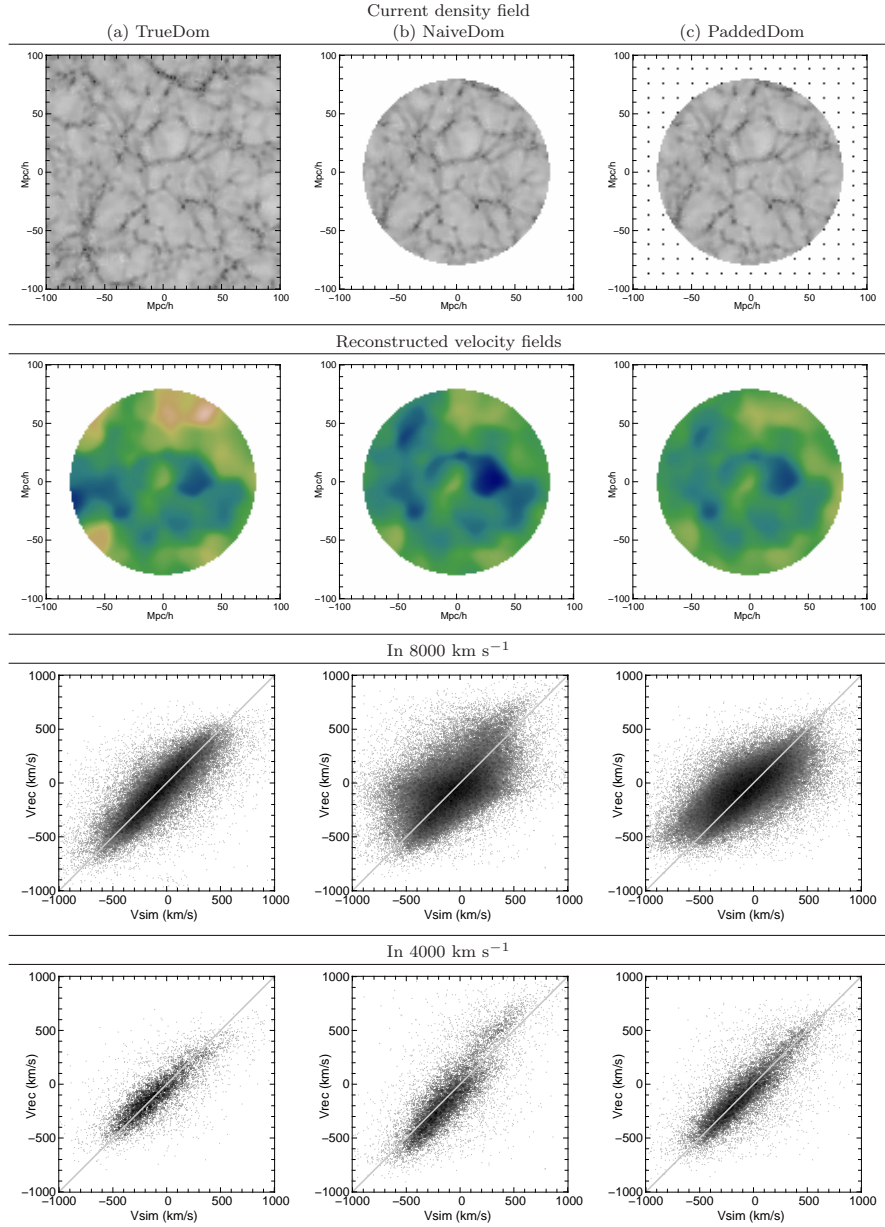
- *PaddedDom* reconstruction is obtained by padding homogeneously the 8k-mock6 catalogue. The padding is chosen such that the final MAK mesh that will be reconstructed is an inhomogeneous cube (as in right panel of the second row of Fig. 15 and 16). The cube must be sufficiently big to absorb density fluctuations present at the boundary of the catalogue (typically a  $20 h^{-1}\text{Mpc}$  buffer zone is needed). With real data, we are bound to assume that the catalogue is totally representative of the whole universe, *i.e.* its effective mean matter density is equal to  $\Omega_m$ .

Fig. 15 shows the result of a *TrueDom*, *NaiveDom* and *PaddedDom* reconstruction applied to 8k-mock6 in the absence of redshift distortion. Fig. 16 gives the same reconstructions when applied to a redshift catalogue. Table 5 summarises the value of the moments of  $P(v_{r,\text{sim}}, v_{r,\text{rec}})$  for different cases. We will now first confront the results of real space reconstructions, and second redshift space reconstructions.

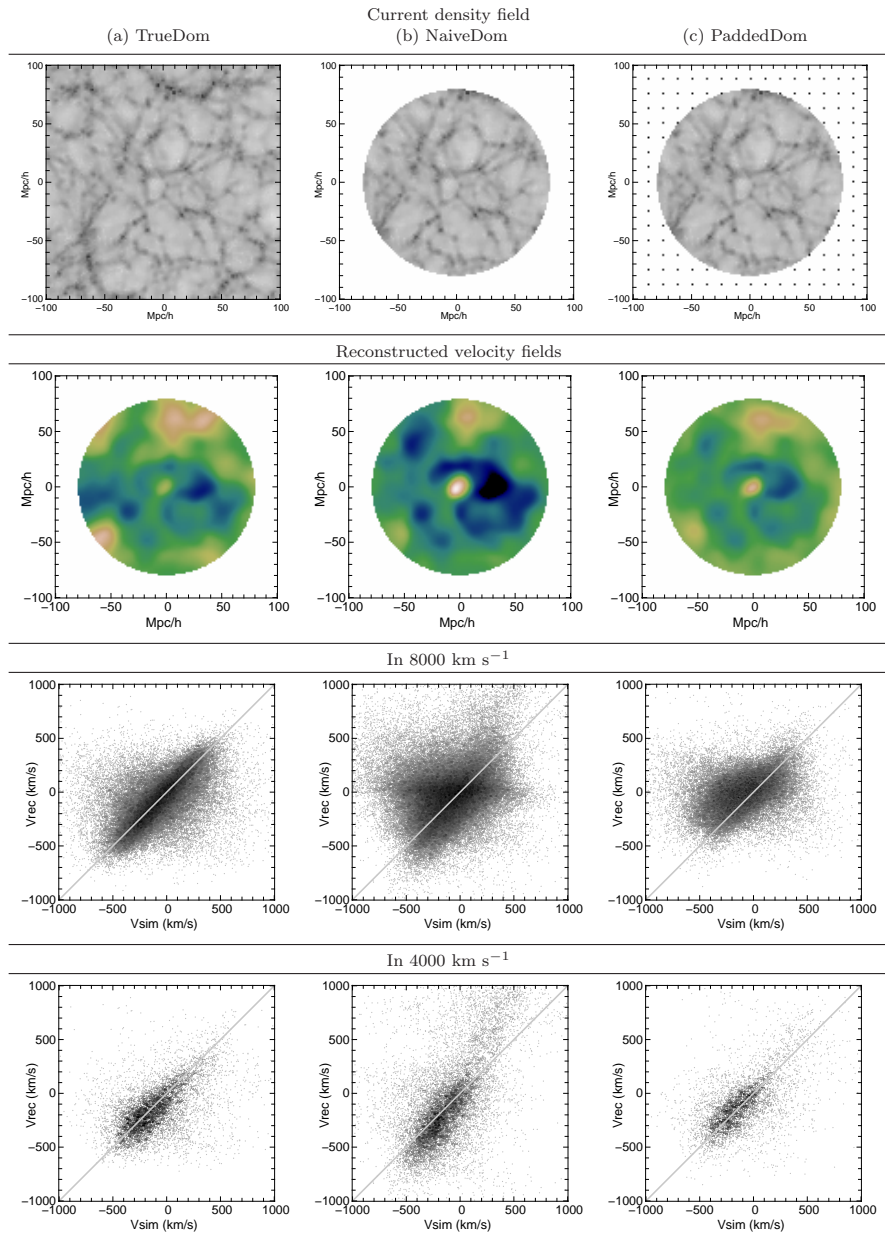
*TrueDom* reconstruction does not yield any significant bias at  $80 h^{-1}\text{Mpc}$ . However, at  $40 h^{-1}\text{Mpc}$ , cosmic variance effects introduce a noticeable systematic error in the direc-

tion of higher  $\Omega_m$  that will be discussed in § 5.3. Compared to *TrueDom*, *NaiveDom* gives good overall results though the central blue region of *TrueDom* turns to dark blue in *NaiveDom*, which would suggest the velocity field is biased. This analysis is confirmed by looking at the bottom scatter plot. The  $\Omega_m$  measurement (Table 5) is underestimated by about 26% even in the central region of the catalogue which is normally less affected by boundary effects. *PaddedDom*, on the other hand, does not yield such a sharp discrepancy in the middle of 8k-mock6, namely in the 4k-mock6 region. Both the bottom scatter plot and the  $\Omega_m$  measurement confirm that the reconstructed velocities are nearly bias-free in the central region. As expected, the velocities in the neighbourhood of the boundaries are completely wrong for the two methods.

Now, the catalogues are cut in redshift space. Redshift distortion biases the velocity distribution of objects on the catalogue boundary: the catalogue receive more infalling objects than outfalling ones. In some cases, one may even find objects seemingly artificially separated from the main volume of the catalogue (they look “disconnected”). In those cases, the hypothesis of convexity is definitely lost for those objects. This problem will enhance boundary problems. The case of *TrueDom* reconstruction has been discussed in § 4. As previously, the peculiar velocities in *NaiveDom* and in *PaddedDom* are largely uncorrelated in the full 8k-mock6 volume (Fig. 16). However, peculiar velocities reconstructed by *NaiveDom* are more strongly overestimated than by using *PaddedDom*’s, as shown in Table 5. For *NaiveDom*, the scatter is plagued by a horizontal alignment in Fig. 16, mid-lower panels, which is a signature of a strong edge effect. This spurious alignment was already present, though much less



**Figure 15.** *Lagrangian domain / without redshift distortions* – This figure summarises the results obtained on reconstructions that have limited information on the Lagrangian domain. The left column illustrates the *TrueDom* reconstruction, the middle column the *NaiveDom* one, and the right column the *PaddedDom* one. For space occupation reasons, the original velocity field given by the simulation is not remembered but can be found in Fig. 11. The *top row* illustrates the three schemes for handling boundary effects on the density field: in the left column one retains information of large scale tidal fields, in the middle column one cuts the catalogue spherically and does a reconstruction on it, in the right column one pads the spherically-cut catalogue with particles homogeneously distributed on a grid. The *second row* gives the reconstructed velocity field in each case, smoothed with a  $5 h^{-1} \text{ Mpc}$  Gaussian window as usual. The color coding is the same as for the other figures, *i.e.* dark blue is  $-1000 \text{ km s}^{-1}$  and white is  $+1000 \text{ km s}^{-1}$ . The *third row* compares the individual (not smoothed) reconstructed and simulated velocities of objects in the 8k-mock6 catalogue. The *fourth row* does the same comparison but objects lying only in the  $4000 \text{ km s}^{-1}$  region of the 8k-mock6 catalogue.



**Figure 16.** *Lagrangian domain / with redshift distortions* – Same as Fig. 15, but for mock catalogues including redshift distortion.

apparent, in the real space case. On the other hand, *PaddedDom* does not present this feature but only a large scatter. We have verified that objects belonging the horizontal alignment are essentially near the  $80 h^{-1} \text{Mpc}$  boundary, contrarily to velocities reconstructed using *PaddedDom* which are more or less uniformly distributed and essentially uncorre-

lated to simulated velocities.<sup>10</sup> This means that *PaddedDom*

<sup>10</sup> This behaviour is expected from an algorithmic point of view. The objects nearby the boundary cannot acquire any displacement using MAK because of the “pressure”/competition of ob-



is at least better at screening edge effects than *NaiveDom* in the sense the errors are more evenly distributed and less systematic. Though impressively low in the last two rows of Table 5, the correlation coefficient  $r$  is actually spoiled by the long tails of the PDF shown in the scatter plots in Fig. 16. Concerning  $\Omega_m$ , *NaiveDom* seems less robust to produce an unbiased estimation than *PaddedDom*. Indeed, looking at Table 5, one may note that the interval delimited by  $s_{\text{med}}$ ,  $s_{\text{min}}$  and  $s_{\text{max}}$  nearly does not contain  $\Omega_m = 0.30$  for *NaiveDom*/Real space/ $40 h^{-1}\text{Mpc}$ , and does not contain it at all for *NaiveDom*/Redshift space. On the contrary,  $\Omega_m = 0.30$  is always selected by the three  $s$  parameters using *PaddedDom* reconstruction. In the rest of this paper, whenever it is needed, we will thus use the *PaddedDom* reconstruction.

### 5.3 Cosmic variance

We generally assume that galaxy catalogues give a fair representation of the whole universe, but of course we have no guarantee that this assumption is correct. Thus, the result of a MAK reconstruction may be affected by inhomogeneities above the catalogue scale. For instance, our galaxy may reside in a particularly extreme region (overdense or underdense), which would produce unusual peculiar velocities. This effect, known as *cosmic variance*, can be investigated by our three original basic mock catalogues: 4k-mock6, 4k-mock7, 4k-mock12 (§ 1). The cosmic variance effect is here further enhanced by the finiteness of the sampled volume. The volume is sufficiently small here to have a non-zero average line-of-sight velocity. On a  $40 h^{-1}\text{Mpc}$  scale, this effect can substantially modify the  $\Omega_m$  measurement (put  $\Omega_{m,\text{mes}}$  in this section) by cutting the  $P(v_{\text{rec}}^s, v_{\text{sim}})$  distribution at an inadequate place.

The results of the reconstruction on these three mock catalogues are given in Fig. 17. In Table 6, we give, for each mock catalogue, the best achievable result (thus highlighting purely the effect of choosing this mock catalogue) and the results one would obtain through observation of this piece of the universe. Unknown Lagrangian domain, redshift distortion and incompleteness effects are added to the considered mock catalogue. The problems of mass-to-light assignment and the zone of avoidance are left apart for the sake of clarity. Their imprint on the velocities should most likely remain the same as we have shown in the corresponding previous sections, *i.e.* biasing for the first and increase of the scatter for the second. Only the cases with the forementioned observational effects are represented in Fig. 17.

Visual inspection of lower scatter plots in Fig. 17 shows that volume finiteness is likely making the  $\Omega_{m,\text{mes}}$  measurement sensitive to the “local”  $\Omega_m$  ( $\Omega_{\text{eff}}$  in the table). This assertion is supported by the estimation of  $s$  and  $\Omega_m$  for *TrueDom* reconstructions given in Table 6. Moreover, experiments conducted with the spherical collapse model show

jects inside the sphere. This problem is further enhanced in redshift space because generally these objects come from outside the sphere and are selected because their infall velocity is high. In *NaiveDom*, they cannot escape from the assumed spherical Lagrangian domain which thus leads to zeroing their velocity. On the other hand, *PaddedDom* is much less strict on the boundary, which leaves the freedom for MAK reconstruction to have a non-zero velocity even for objects on the boundary of the catalogue.

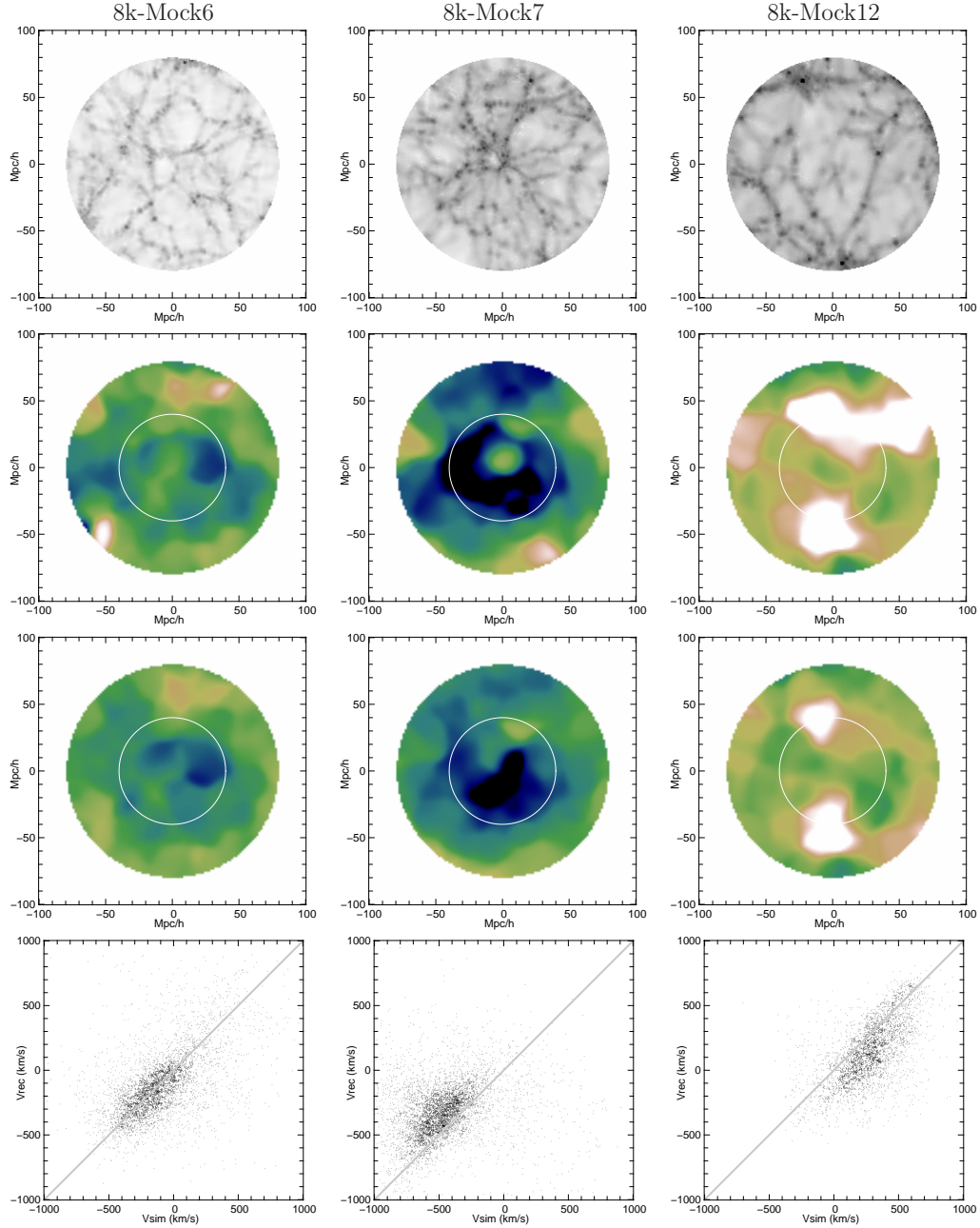
that  $\Omega_{m,\text{mes}}$  is indeed a weighted average between  $\Omega_{\text{eff}}$  and  $\Omega_m$ .

More specifically, reconstructed velocities in *4k-mock7* (including observational effects) are apparently giving the  $\Omega_m$  of the simulation but they present a large scatter rendering the slope estimation dubious. Indeed, doing the same reconstruction but without observational effects give a measured  $\Omega_{m,\text{mes}} = 0.40$ , which is the exact average between the simulation  $\Omega_{m,\text{simu}} = 0.30$  and  $\Omega_{\text{eff}} = 0.50$ .<sup>11</sup> The aforementioned scatter is expected for this mock catalogue: the velocity field is badly reconstructed near the observer in that case (middle panels) because the local cosmic flow is higher than usual ( $\sim 1000 \text{ km s}^{-1}$ ) and the non-linearities are stronger. Thus the convexity of the problem is lost on an extended region around the observer when the reconstruction is conducted in redshift space (see § 4). A particularly salient mis-reconstruction is given by the outflowing “bubble” at the center which disappears in the reconstructed velocity field. The size of the affected region is about  $20 h^{-1}\text{Mpc}$  around the observer in 4k-mock7 and thus limits the number of objects having both good reconstructed and observable peculiar velocities.

In an opposite way, velocities in *4k-mock12* are reconstructed with a better correlation, as shown by Table 6, but  $\Omega_m$  measurement is strongly weighted toward  $\Omega_{\text{eff}}$ . These two “features” are largely due to the huge central void. First, MAK reconstruction and Zel’dovich approximation are known to work better in low density regions and being centered on a void results in inhibiting blueshift distortion as galaxies are principally going away from the observer, rendering the reconstruction problem convex in Eq. (8). Second, the low density region largely affects the statistical velocity distribution, which in this case leads to a measured  $\Omega_{m,\text{mes}}$  weighted more strongly towards the  $\Omega_{\text{eff}}$  of 4k-mock12.<sup>12</sup> This leads us to a  $\Omega_{m,\text{mes}}$  that is nearer  $\Omega_{m,\text{simu}}$  in 4k-mock12 than the mean matter density of the whole simulation. The volume finiteness also produces an apparent offset between reconstructed velocities and measured ones. This is expected as doing a statistical analysis on a finite volume catalogue must introduce a selection bias effect. We have indeed checked that the point set  $\{(v_{r,i}, \psi_{r,i}^s)\}$ , obtained through a MAK reconstruction applied on 4k-mock12, is a subset of the corresponding set built from a reconstruction on 8k-mock12. Looking at our “standard” *4k-mock6*, one can note that the simulated velocity distribution is generally more symmetric according to the null velocity than for the two other mock catalogues, with no visual bias while comparing reconstructed velocity to simulated velocity. This supports the initial assertion linking  $\Omega_{m,\text{mes}}$  to  $(\Omega_{m,\text{simu}}, \Omega_{\text{eff}})$  and the asymmetric distribution of velocities. Potentially, one could recover the true  $\Omega_m$  of the Universe (or here the simulation) from the measured velocities of any catalogues by predicting how the velocity distribution asymmetry is linked to local density contrast. However, the simplest, and more robust solution, would still

<sup>11</sup> Spherical collapse rather predicts  $\Omega_{m,\text{mes}} = 0.35$  for the same setup.

<sup>12</sup> The spherical collapse model would predict a measured  $\Omega_{m,\text{mes}} = 0.26$  and this is in good agreement with the value measured when no observational effects are injected in the mock catalogue.



**Figure 17.** *Cosmic variance* – This figure gives a visual comparison of the three mock catalogues used to study cosmic variance effects. Top panels: Adaptively smoothed density fields of the considered mock catalogues. In each case, we have represented the central thin slice that contains the observer. Second row: Simulated velocity field, after smoothing with a  $5 h^{-1} \text{Mpc}$  Gaussian window. The white circle gives the limit of the  $40 h^{-1} \text{Mpc}$  volume. Third row: Same as second row, but for the reconstructed velocity field. Fourth row: Comparison between reconstructed and simulated peculiar velocities.

**Table 5.** *Lagrangian volume* – Residual error after the correction. Description for some columns is given in the caption of Table 1. “Radius” gives the spatial size of the sphere on which the velocity-velocity comparison is conducted. “Reconstruction type” indicates the type of Lagrangian domain reconstruction and whether it is mixed with redshift distortion effect. Details on the meaning of each name are given in § 5.2.

Reconstruction type	Radius ( $h^{-1}\text{Mpc}$ )	Velocities			$\Omega_m$ (s)	$\Omega_m$ ( $\mathcal{L}_{\min}$ )	$\Omega_m$ ( $\mathcal{L}_{\max}$ )	$\Omega_m$ ( $1.5\sigma, s_{\text{med}}$ )	$\Omega_m$ ( $1.5\sigma, s_{\min}$ )	$\Omega_m$ ( $1.5\sigma, s_{\max}$ )
		s	r	$\sigma$						
TrueDom / Real space	80	0.91	0.77	0.66	0.35	0.28	0.31	0.27	0.233	0.32
	40	0.80	0.76	0.65	0.45	0.28	0.38	0.35	0.28	0.43
NaiveDom / Real space	80	0.87	0.52	0.92	0.38	0.20	0.28	0.42	0.20	0.87
	40	1.11	0.77	0.73	0.25	0.20	0.24	0.244	0.19	0.31
PaddedDom / Real space	80	0.73	0.65	0.77	0.53	0.36	0.48	0.45	0.27	0.75
	40	0.91	0.77	0.64	0.35	0.28	0.34	0.32	0.26	0.38
NaiveDom / Redshift space	40	1.49	0.51	1.31	0.11	0.15	0.26	0.20	0.12	0.37
PaddedDom / Redshift space	40	0.93	0.53	0.94	0.36	0.18	0.34	0.38	0.20	0.79

**Table 6.** *Cosmic variance* – Summary of measurements conducted on the three mock catalogues. The reconstruction is either conducted on the basic catalogue without any observational effect besides cosmic variance (labelled *Original*), or on the same catalogue but affected by redshift distortion, incompleteness and for which the Lagrangian domain is determined using *PaddedDom* (reconstruction labelled *Full*). The description of the other columns are given Table 1.

Catalog	Reconstruction type	Velocity field			$\Omega_m$ (s)	$\Omega_m$ ( $\mathcal{L}_{\min}$ )	$\Omega_m$ ( $\mathcal{L}_{\max}$ )	$\Omega_m$ ( $1.5\sigma, s_{\text{med}}$ )	$\Omega_m$ ( $1.5\sigma, s_{\min}$ )	$\Omega_m$ ( $1.5\sigma, s_{\max}$ )
		s	r	$\sigma$						
4k-mock6 ( $\Omega_{\text{eff}} = 0.35$ )	Original	0.80	0.76	0.65	0.313	0.28	0.38	0.35	0.28	0.43
	Full	0.94	0.50	0.96	0.35	0.13	0.31	0.31	0.16	0.70
4k-mock7 ( $\Omega_{\text{eff}} = 0.5$ )	Original	0.70	0.67	0.76	0.57	0.39	0.47	0.40	0.33	0.48
	Full	0.88	0.11	1.33	0.43	0.41	1.62	0.30	0.09	1.29
4k-mock12 ( $\Omega_{\text{eff}} = 0.19$ )	Original	1.12	0.81	0.66	0.24	0.235	0.27	0.24	0.22	0.26
	Full	1.08	0.58	1.11	0.24	0.29	0.62	0.15	0.08	0.31

be to extend the depth of current catalogues to reach a volume where velocities are normally distributed.

From a prediction point of view, comparing visually the velocity fields inside the white circles show that, if we know  $\Omega_m$ , we reconstruct plausible velocity fields for the three mock catalogues. Outside the white circles, the reconstructed velocity field is nearly completely uncorrelated compared to the simulated one as we have discussed in the previous section. It must be noted that the velocity field goes smoothly to zero (green colour) on the edge of all mock catalogues: this is an expected side effect of the homogeneous padding which tends to smooth out any fluctuation on the edge (velocity and density field).

## 6 VELOCITY MEASUREMENT ERRORS

### 6.1 The need for a likelihood analysis ?

All the effects already described in this paper are present in a redshift catalogue. Though we expect most of the observational biases should be independent, some of them may correlate and give worse systematic errors. We present in Fig. 18 the progressive deterioration of the velocity-velocity comparison for *4k-mock6* based on a reconstruction conducted on *8k-mock6*. The effects are piled up from left to right. The  $\Omega_m$  measurements for the  $1.5\sigma$  method are indi-

cated below each panel. The obvious conclusion is that the measurements are progressively affected but that no extra correlated error seems to happen when mixing the effects. Another fortunate event is that bias seems to counterbalance themselves to give in the end a nearly unbiased result (last but one panel). Going from *TrueDom*/Real to Redshift tends to decrease  $\Omega_m$  as has been seen previously. On the contrary, injecting incompleteness pushes the measurement to higher  $\Omega_m$  as we have noticed in § 3.3. The  $1.5\sigma$  method seems to give the right  $\Omega_m$  value in all cases, which means that we should be able to use it on galaxy catalogues provided we have sufficient precision on velocity measurements. However, looking at the last panel (bottom right) of Fig. 18 shows that injecting random velocity measurement errors (here we introduced an optimistic error of 8% of the distance to the object, corresponding to an error on distance modula of  $\sigma_\mu = 0.17$ ), renders slope estimation much more difficult. In that case, the measured  $\Omega_m$  is severely biased. This is expected as the  $1.5\sigma$  method relies mostly on the central part of the scatter, which in turn is the one that is the most affected by random errors. This leads to a circularization of the  $1.5\sigma$  isocontour and thus a completely wrong estimation of the slope. On the other hand, looking at the global structure of the scatter shows that the right slope is still hidden in the data, but one should then take into account the tails of the distribution. This last test shows the



## Observational biases in Lagrangian reconstructions of cosmic velocity fields 21

limit of a direct velocity-velocity comparison in real cases. It might be possible to recover the original distribution of the scatter by deconvolving from the noise. However, it seems to be a difficult operation and we prefer to first try a maximum likelihood approach. Its main advantage would be to work using distances, thus rendering the error in measurements more tractable.

## 6.2 Maximum likelihood analysis

Observations of galaxies first give us access to their distances and not their peculiar velocities. A method based on distances to make a comparison between a model and observations is potentially less sensitive to distance measurement errors. Indeed, by comparing directly distances, one has a small relative error on each measurement instead of a huge one when peculiar velocities are considered. Below, we discuss galaxy selection bias and zero-point calibration errors in distance measurements while keeping the notation of Strauss & Willick (1995).

*Presentation of the Bayesian chain* – For the Tully-Fisher (TF) relation, one makes an estimate of the absolute magnitude of a galaxy as a function of its linewidth: the slope between the two quantities can be biased because the sample is limited in magnitude (Strauss & Willick 1995). This effect which is known as *selection bias* is purely statistical and if not correctly taken into account can lead to large systematic errors. Using these absolute magnitudes, occasionally combined to form groups of galaxies, and the apparent magnitudes of the same group, one builds the distance modulus

$$\mu(r) = m(r) - M = 5 \log_{10} \left( \frac{r}{10 \text{ pc}} \right) \quad (10)$$

with  $r$  the distance of the considered object (group of galaxies or galaxy). In addition to the forementioned statistical bias, peculiar velocity obtained from redshift positions through a Lagrangian reconstruction, here MAK, are sometimes very noisy, as shown in Fig. 18. Another more subtle effect is introduced by the Gaussian distribution of our velocity sample that we are going to analyze. We need to take care of this “selection bias” to avoid being spoiled by eventual large reconstruction errors present for objects with a high velocity. Thus we need a Bayesian approach to account for all these statistical effects.

In principle, the likelihood function gives a probability for the data, *i.e.* here redshift positions  $\mathfrak{Z} = \{\mathbf{z}_i\}$ , with  $i$  running from 1 to  $N$ , and distance moduli  $\mathfrak{M} = \{\mu_i\}$ , assuming some model described by the vector parameter  $p$ . Additionally we assume that we have an estimation of measurement errors on  $\mathfrak{M}$  through the set  $\mathfrak{S}$ . The exact description of  $\mathfrak{S}$  will be given in the next paragraph. Typically errors on redshift measurements are of the order of 50-60 km s<sup>-1</sup>. This means that we can consider them as negligible if we consider objects farther than  $R_z = 6 - 10 \text{ h}^{-1} \text{ Mpc}$ . The volume enclosed by the sphere of radius  $R_z$  is, in any case, also poorly reconstructed because of the singularity introduced by redshift distortions near the observer (§ 4). In the following analysis, we will consider redshift measurements as negligible by avoiding the objects located at less than  $10 \text{ h}^{-1} \text{ Mpc}$

from the observer, thus we have:<sup>13</sup>

$$P(\mathfrak{M}, \mathfrak{Z} | p, \mathfrak{S}) \propto P(\mathfrak{M} | \mathfrak{Z}, p, \mathfrak{S}) = \mathcal{L}(p) \quad (11)$$

The end of this section is devoted to computing the right hand part of this equation. To achieve this, we will decompose the probability into small pieces:

$$P(\mathfrak{M} | \mathfrak{Z}, \mathfrak{S}, p) = \iiint_{\mathfrak{M}_r, \mathcal{V}, D} P(D) P(\mathfrak{M} | \mathfrak{M}_r, \mathfrak{S}, D, p) \times P(\mathfrak{M}_r | \mathcal{V}, \mathfrak{Z}, p) P(\mathcal{V} | \mathfrak{Z}, p) d\mathfrak{M}_r d\mathcal{V} dD \quad (12)$$

with  $\mathfrak{M}_r = \{\mu_{1,r}, \dots, \mu_{N,r}\}$  representing the “true” distance moduli, with  $\mu_{i,r} \in [-\infty, +\infty]$  and  $\mathcal{V} = \{v_1, \dots, v_n\}$  the “true” object peculiar velocities.  $P(\mathfrak{M} | \mathfrak{M}_r, \mathfrak{S}, p)$  is the probability of measuring the set of distance moduli  $\mathfrak{M}$  given that the real set of distance moduli is  $\mathfrak{M}_r$  and the expected error on the measurement is given by  $\mathfrak{S}$ .  $P(\mathfrak{M}_r | \mathcal{V}, p)$  is the probability of obtaining the set of distance moduli  $\mathfrak{M}_r$  given the reconstructed velocities  $\mathcal{V}$ .  $P(\mathcal{V} | \mathfrak{Z}, p)$  is the probability the velocities are well reconstructed from the redshift data  $\mathfrak{Z}$ . The probability  $P(D)$  is going to be introduced in the last paragraph to account for uncertainty in the calibration of the Tully-Fisher relation. All those probabilities are computed assuming the model parameters  $p$ . We will establish the likelihood function  $\mathcal{L}(p)$  in three steps:

- First, the error distributions linked to observations are considered to get an unbiased distance estimator for groups. This analysis yields the probability  $P(\mu_i | \mu, \sigma_{\mu,i}, p)$ .
- Second, the errors on reconstructed velocities are considered to compute  $P(v | \mathfrak{Z}, p)$ .
- Last, the two analyses are merged as given above to produce the likelihood function which gives the posterior distribution of  $\beta$  and the Hubble constant  $H$ .

A picture of the above Bayesian chain is given in Fig. 19.

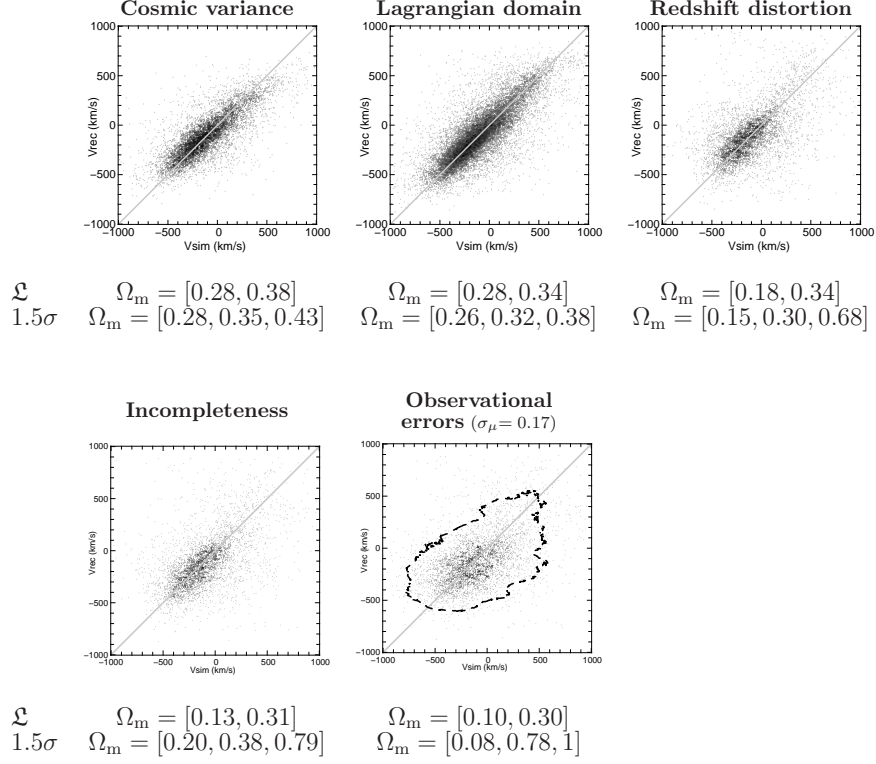
*Distance modulus error distribution* – To establish the likelihood function comparing the measured distance to the reconstructed velocity field, we assume the distance catalogues are obtained using the inverse TF relation (Shaya et al. 1995),

$$\eta^0(M) = -e(M + D), \quad (13)$$

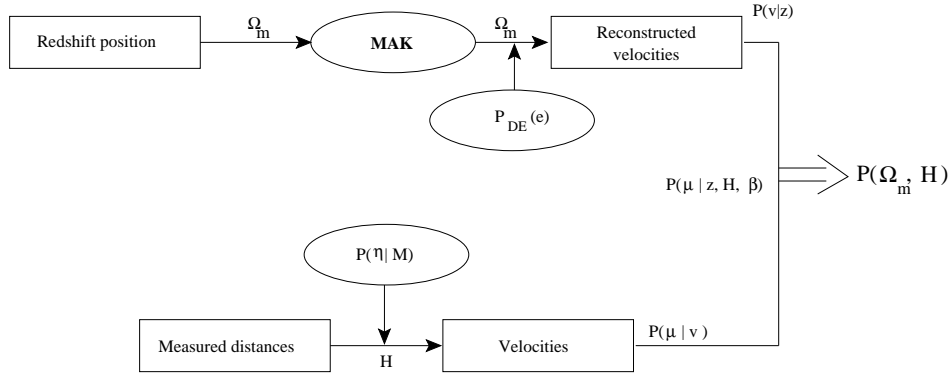
where  $M$  is the absolute magnitude of the considered galaxy,  $\eta^0(M)$  is its predicted linewidth,  $e$  is the slope, and  $D$  is the zero point calibration (the latter two are assumed to be known exactly). It is known that inverse TF is less sensitive to the selection bias as compared to forward TF (Strauss & Willick 1995). Observational data show that the differences between the predicted linewidth  $\eta^0(M)$  and the measured linewidth  $\eta$  for an object of absolute magnitude  $M$  are Gaussian distributed<sup>14</sup> (Pizagno et al. 2006; Tully & Pierce

<sup>13</sup> Though it is in theory possible to avoid this hypothesis, it is in practice highly difficult for computational reason as one would need to run several MAK reconstructions to evaluate the extra integral that would be needed in Eq. (11).

<sup>14</sup> In fact, in writing Eq.(14), two effects are mixed: the error on the measurement of linewidth, which may reach 10% because of the uncertainty in galaxy inclination correction, and the intrinsic modeling errors of the TF relation itself.



**Figure 18.** This figure gives the evolution of the scatter distribution and of the measurement of  $\Omega_m$  using it while more and more observational effects are added to 8k-mock6 catalogue. All measurements of  $\Omega_m$  are given between brackets and are sorted as follows. For measurements obtained through the likelihood analysis, labelled by  $\mathcal{L}$ , the first number corresponds to  $\mathcal{L}_{\text{min}}$  and the second to  $\mathcal{L}_{\text{max}}$ . For measurements obtained using the 1.5 $\sigma$  method, the first number corresponds to  $s_{\text{min}}$ , then  $s_{\text{med}}$  and finally  $s_{\text{max}}$ . The last (lower right) panel uses the full likelihood function of Eq. (31). All others use a restricted likelihood analysis with  $\sigma_0/e = 0$ , which is nearly equivalent to using Eq.(30) for each  $(v_{r,i}, \psi_{r,i})$  pair. The 1.5 $\sigma$  isocontour has been plotted with a thick dashed line in the last panel.



**Figure 19.** *Maximum likelihood analysis* – This sketch illustrates the bayesian chain used to establish the likelihood function. The input data are located on the left and the output posterior distribution  $P(\beta, H)$  on the right.

2000). Thus, the probability of measuring the linewidth  $\eta$ , given that the object has an absolute magnitude  $M$ , and assuming that the TF relation  $\eta^0(M)$  is known, is

$$P(\eta|M, e, D) = \frac{1}{\sqrt{2\pi}\sigma_\eta(M)} e^{-\frac{(\eta - \eta^0(M))^2}{2\sigma_\eta^2(M)}} \quad (14)$$

with  $\sigma_\eta(M)$  the linewidth estimation error for the absolute magnitude  $M$ . Distance catalogues are composed of estimated distance moduli  $\mu_e$  from the inverse TF relation. These estimated distance moduli are built from the statistics on a single group. Therefore, the joint probability of having a galaxy in a group with both a linewidth  $\eta$  and an absolute magnitude  $M$ , assuming the TF relation  $\eta^0(M)$ , is:

$$P(\eta, M|e, D) = F(M)P(\eta|M, e, D), \quad (15)$$

where  $F(M)$  is the normalized absolute luminosity function of the group.<sup>15</sup>

The estimator for the distance modulus is given by:

$$\mu_e = m - M_e(\eta) = M + \mu_0(r) + D' + \frac{\eta}{e'}, \quad (16)$$

where  $D'$  and  $e'$  are the estimated inverse TF parameters of Eq. (13) and  $\mu_0(r) = 5 \log(\frac{r}{10 \text{ pc}})$  the true distance modulus of the considered group. The conditional probability that the estimated distance modulus for the group is  $\mu$ , assuming that the estimated Tully-Fisher parameters are  $e'$  and  $D'$  and that the real parameters for this group are  $e$  and  $D$ , can be written as

$$\begin{aligned} P(\mu|\mu_0(r), e, e', D, D') &= \langle \delta_D (\mu - \mu_e) \rangle_{\text{group}} \\ &= \left\langle \delta_D \left( \mu - \mu_0(r) + M + D' + \frac{\eta}{e'} \right) \right\rangle_{\text{group}} \\ &= \int_M \frac{e' F(M)}{\sqrt{2\pi}\sigma_\eta(M)} e^{-\frac{(e'(\mu_0(r) - \mu) + e' D' - e D + (e' - e) M)^2}{2\sigma_\eta^2(M)}} dM. \end{aligned} \quad (17)$$

While working with the inverse TF relation, one can assume that the slope  $e'$  is completely determined and  $e' = e$ . Since the observed  $\sigma_\eta(M)$  varies little with  $M$ , it is chosen to be equal to a constant  $\sigma_0$ . The previous probability reduces to

$$P(\mu|\mu_0(r), e, D, D', \sigma_0) = \frac{e}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\mu - \mu_0(r) + D' - D)^2}{2\sigma_0^2}}. \quad (18)$$

Though the slope  $e'$  is well determined, the zero-point calibration  $D$  may still be affected by non-negligible errors.<sup>16</sup> The set describing errors on distance moduli is thus  $\mathfrak{S} = \{\sigma_{0,1}/e, \dots, \sigma_{0,N}/e\} = \{\sigma_{\mu,1}, \dots, \sigma_{\mu,N}\}$ . The error on this calibration will affect the distances globally. As a first approximation we model the error on the zero point by a Gaussian centered on  $D$  with a standard deviation of  $\sigma_D$ .

*Linking distance modulus to velocity* – The second probability function in Eq. (12) is  $P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p)$ , which is actually a distribution linking the velocities and redshifts to

distance modulus. This principally corresponds to a change of variable and we give directly the expression of it, which is inspired by Eq. (1):

$$P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p) = \prod_{i=1}^N H 10^{\mu_{r,i}/5} \delta_D \left( z_i - v_i - 10 \text{ pc} \times H 10^{\mu_{r,i}/5} \right) \quad (19)$$

*Reconstructed velocity distribution* – We are now going to establish the expression of  $P(v|\mathfrak{Z}, p)$  with the vector of parameters of our chosen model  $p = (H, \beta, B_v, \sigma_v, \gamma_*, e) - \sigma_v$  and  $\gamma_*$  are going to be introduced in the next immediate paragraphs. One may decompose  $P(v|\mathfrak{Z}, p)$  that way

$$P(v|\mathfrak{Z}, p) = \int_{\mathfrak{P}} P(v|\mathfrak{P}, p) P(\mathfrak{P}|\mathfrak{Z}, p) d\mathfrak{P}, \quad (20)$$

with  $\mathfrak{P} = \{\psi_{r,i}\}$  the reconstructed displacements. As MAK reconstruction is deterministic once  $\beta$  has been assumed (§ 4), the second probability distribution is simply given in our case by

$$P(\mathfrak{P}|\mathfrak{Z}, p) = \prod_{i=1}^n \delta_D (\psi_{r,i} - \psi_i(\mathfrak{Z}, \beta)) \quad (21)$$

with  $\psi_i$  representing the MAK reconstructed displacement of the  $i$ -th object, being a function of all redshift coordinates and  $\beta$ . Thus, studying  $P(\mathcal{V}|\mathfrak{Z}, p)$  reduces to examine  $P(\mathcal{V}|\mathfrak{P}(p_0), p')$ , with  $p' = (H, \beta', B_v, \sigma_v, \gamma_*)$ ,  $p_0 = (H, \beta_0, B_v, \sigma_v, \gamma_*)$ ,  $\beta_0$  being the assumed growth factor to compute the set  $\mathfrak{P}(p_0)$  using the redshift reconstruction.  $P(\mathcal{V}|\mathfrak{Z}, p)$  and  $P(\mathcal{V}|\mathfrak{P}(p_0), p')$  equalizes only if  $p = p' = p_0$ . Thus one needs a several redshift reconstructions to build the probability function  $P(\mathcal{V}|\mathfrak{Z}, p)$ . Working with the intermediary set  $\mathfrak{P}$  is easier than with  $\mathfrak{Z}$ , we thus put the reduced likelihood function:

$$\begin{aligned} \mathcal{L}'_{\beta_0}(p') &= \iint_{\mathfrak{M}_r, \mathcal{V}} P(\mathfrak{M}_r|\mathfrak{S}, \mathfrak{S}) P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p) P(\mathcal{V}|\mathfrak{P}(p_0), p') d\mathfrak{M}_r d\mathcal{V} \end{aligned} \quad (22)$$

and we are going to establish the expression of the elementary probability function  $P(v_r|\psi_r, p)$  which will yield

$$P(\mathcal{V}|\mathfrak{P}(p_0), p) = \prod_{i=1}^N P(v_{r,i}|\psi_{r,i}, p) \quad (23)$$

assuming statistical independance of all  $\{v_{r,i}, \psi_{r,i}\}$  duets, and that  $\mathfrak{P}(p_0)$  is obtained using a redshift reconstruction for which  $\beta = \beta_0$ .  $\mathcal{L}'$  may be written in a factorized way:

$$\begin{aligned} \mathcal{L}'_{\beta_0}(p') &= \prod_{i=1}^N \iint_{\mu_r, v_r} P(\mu_i|\mu_r, \sigma_{0,i}/e) P(\mu_r|v_r, z_i, p) P(v_r|\psi_{r,i}, p) d\mu_r dv_r. \end{aligned} \quad (24)$$

The computation of  $\mathcal{L}'$  is clearly helped using this factorized form. We may now concentrate on the third probability function of the above equation.

As has been established in § 2, the distribution of errors on the reconstructed velocity field is the Lorentzian

$$P_{DE}(e_{\psi v}) \propto \frac{1}{1 + \left( \frac{e_{\psi v}}{B_v} \right)^2}, \quad (25)$$

<sup>15</sup> Note that the selection function is assumed to be independent of  $\eta$  and is hence absorbed in  $F(M)$ .  $F(M)$  corresponds to  $\Phi(M)S(M, \eta)$  in Strauss & Willick (1995) notation, e.g. eq. (188).

<sup>16</sup> The latest calibration is given in Tully et al. (2007).

where  $B_v = 86 \text{ km s}^{-1}$  (redshift reconstruction), with  $e_{\psi v}$  the distance between the reconstructed velocity  $\beta\Psi_r$  and the true velocity  $v_r$ . This formulation is different from saying that the reconstructed velocity is affected by error when compared to the true velocity, and permits some errors in the MAK reconstructed displacement field. As has been seen in § 5.3, the reconstructed velocities may also contain an extra offset that needs to be removed while measuring  $\beta$ . The error distance  $e_{\psi v}$  is thus

$$e_{\psi v} = \alpha_* v_r - \beta_* \Psi_r + \gamma_* \quad (26)$$

with

$$\alpha_*^2 + \beta_*^2 = 1 \text{ and } \beta = \beta_*/\alpha_* \quad (27)$$

and  $\gamma_*$  to account for a potential spurious offset in reconstructed velocities. From linear theory (Peebles 1980), we know that the line-of-sight component of the velocity field must be distributed like a Gaussian function. We now assume that the absolute probability for an object to have a velocity  $v$  is given by a Gaussian distribution:

$$P_{\text{vel}}(v|p) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{v^2}{2\sigma_v^2}}. \quad (28)$$

It must be noted that it is likely that the observational data does not encompass a sufficiently large volume so that measured velocities follow this law. Moreover, this prior is of some importance when we have to deal with highly scattered data. The shortcomings of such an approach will be discussed in the next section. One can recover the standard uniform prior on velocities by taking the limit  $\sigma_v \rightarrow +\infty$  in the next equations. Assuming  $e_{\psi v}$ , as a random variable, is independent of  $v_r$  and these two quantities are themselves statistically independent from  $\beta$  and  $\gamma_*$ , we may now write the joint probability of reconstructing  $\psi_r$ , having a true velocity  $v_r$ :

$$\begin{aligned} P(v_r, \psi_r, \beta, \gamma_* | B_v, \sigma_v) &= \beta_* P_{\text{DE}}(e_{\psi v} | B_v, \sigma_v) \times \\ &\quad P(\beta, \gamma_* | B_v, \sigma_v) P(v_r | B_v, \sigma_v) P(\psi_r | B_v, \sigma_v) \\ &= \beta_* C(B_v, \sigma_v) P(\beta, \gamma_* | B_v, \sigma_v) \frac{P(\psi_r | B_v, \sigma_v) e^{-\frac{v_r^2}{2\sigma_v^2}}}{1 + \left( \frac{\beta_* \psi_r - \alpha_* v_r + \gamma_*}{B_v} \right)^2}, \end{aligned} \quad (29)$$

where  $C$  is a function eventually depending on  $B_v$  and  $\sigma_v$ . The conditional probability that the true velocity is  $v_r$  given the reconstructed displacement  $\psi_r$  is now exactly

$$P(v_r | \psi_r, p) = \frac{e^{-\frac{v_r^2}{2\sigma_v^2}} \left( 1 + \left( \frac{\beta_* \psi_r - \alpha_* v_r + \gamma_*}{B_v} \right)^2 \right)^{-1}}{\int_{v=-\infty}^{+\infty} e^{-\frac{v^2}{2\sigma_v^2}} \left( 1 + \left( \frac{\beta_* \psi_r - \alpha_* v + \gamma_*}{B_v} \right)^2 \right)^{-1} dv} dv \quad (30)$$

The denominator of the right hand part of this equation must be computed numerically.<sup>17</sup> It can be shown that, in the limit  $\sigma_v \rightarrow +\infty$ ,  $P(v_r | \psi_r, \beta_*, \gamma_*)$  reverts to a pure Lorentzian form.

<sup>17</sup> This function is known as a Voigt profile.

*Merging the probability distributions* – We may now establish the “elementary” conditional probability for an object  $i$  to get a measured distance  $\mu_i$  given that its reconstructed displacement is  $\psi_{r,i}$ , its redshift is  $z_i$ , the error on the linewidth measurement is  $\sigma_{0,i}$  and the model parameters are  $p'$  in the notation of this section:

$$\begin{aligned} P(\mu_i | \psi_{r,i}(p_0), z_i, \sigma_{0,i}, D', p') &= \iint_{v, \mu_r} P(\mu_i | \mu_r, \sigma_{0,i}, D', p') P(\mu_r | v, z_i, p') \\ &\quad \times P(v | \psi_{r,i}(p_0), p') d\mu_r dv \\ &\propto \frac{eH}{\sigma_{0,i}} \int_{\mu_r} 10^{\mu_r/5} e^{-\frac{\mu_r^2}{2\sigma_{0,i}^2}} (\mu_r - \mu_i + D - D')^2 \\ &\quad \times P\left(v = (z_i - 10 \text{ pc} \times H 10^{\mu_r/5}) | \psi_{r,i}(p_0), p'\right) d\mu_r, \end{aligned} \quad (31)$$

with  $p' = (H, \beta', B_v, \sigma_v, \gamma_*)$ ,  $\psi_{r,i}(p_0)$  being computed assuming the parameters  $p_0$ . Looking closely at this probability, one may notice that changing  $D \rightarrow D' = D + \Delta$  is equivalent to changing  $H \rightarrow H' = H \exp(\Delta/5)$ . Thus the uncertainty in the zero point calibration translates only in an uncertainty on  $H$  and not on the parameters of the model.

We may now write the full formal expression of  $\mathcal{L}'(p)$ , as already sketched in Eq. (22). As specified in the discussion we take  $P(D')$  to be a Gaussian distribution centered on  $D$  and with a standard deviation  $\sigma_D$ . Now we may replace and get:

$$\begin{aligned} \mathcal{L}'_{\beta_0}(p') &\propto \int_{D'=-\infty}^{\infty} e^{-\frac{(D-D')^2}{2\sigma_D^2}} \prod_i P(\mu_i | \psi_{r,i}(p_0), z_i, \sigma_{0,i}, e, p') dD, \end{aligned} \quad (32)$$

with  $i$  running on objects of the catalogue. Assuming a uniform prior on  $\beta$ ,  $H$  and  $\gamma_*$  and taking care of the relation between  $\mathcal{L}'_{\beta}$  and  $\mathcal{L}$  as mentioned above, the Bayes theorem permits us to write

$$P(H, \beta, \gamma_* | \mathfrak{M}, \mathfrak{S}, \mathfrak{Z}, e, D', B_v, \sigma_v) \propto \mathcal{L}(p) = \mathcal{L}'_{\beta}(p). \quad (33)$$

We now have access to the posterior distribution of  $(H, \beta, \gamma_*)$ .

### 6.3 Results

The results of measuring  $\Omega_m$  using the maximum likelihood estimator are presented in the tables using the label  $\mathcal{L}$ .

Except in the case where we consider observational errors, we use a simplified version of  $\mathcal{L}$  by taking  $\sigma_{0,i} = 0$ . While it would have been natural to find the maximum of the likelihood for *all* parameters (including  $B_v$ ,  $\sigma_v$ ,  $\gamma_*$ ), we quickly noticed that it was leading to unacceptably biased measurements and to an unnecessary increase of the parameter space. Moreover, the results quite strongly depends on  $\sigma_v$  and  $\gamma_*$ , especially when the reconstruction noise becomes high as in redshift reconstructions (see Appendix D). We thus propose to discuss the values obtained by setting  $\gamma_* = 0$ ,  $B_v = 90 \text{ km s}^{-1}$  and choosing two values for  $\sigma_v$ . First, we use linear theory to predict the average velocity dispersion of haloes in the universe, this leads to take  $\sigma_v = 326 \text{ km s}^{-1}$  (the  $\Omega_m$  measured that way is labelled

$\mathcal{L}_{\max}$ ). Second,  $\sigma_v = +\infty$  is used to check the influence of recovering a uniform prior on the velocity distribution (labelled  $\mathcal{L}_{\min}$ , respectively).

By looking at all tables of this paper, we noticed that the difference between the two measured  $\Omega_m$  is mostly following the interval defined by  $s_{\min}$  and  $s_{\max}$ . We were expecting such a behaviour ( $\sigma_v$  is more or less controlling the statistical bias of the likelihood function) but not that it would so clearly follow the other method. The more the scatter is important, the more the measurement becomes imprecise as expected. It must however be noted that on average the measure  $\mathcal{L}_{\max}$  suffers less systematic bias than  $\mathcal{L}_{\min}$ . This behaviour is supported by the tests conducted in Appendix D.

The seemingly well estimated  $\Omega_m$  in the lower right panel of Fig. 18 has been computed using the full likelihood analysis. Actually, compared to the  $1.5\sigma$  method for which the measured slope is undefined,  $\mathcal{L}_{\min}$  and  $\mathcal{L}_{\max}$  are basically the same as when no observational errors are introduced.

The correction based on a Gaussian velocity distribution assumption, cannot be entirely trusted for 4k-mock7 and 4k-mock12. As one may note in Fig. 17, the velocity distribution is highly non-Gaussian in these cases. This renders incorrect our distribution modeling in § 6.2. Looking at Table 6, we note that though the measurements on “Original” reconstruction is not strongly affected, we cannot say the same thing using data obtained from “Full” reconstruction. In the first case, the noise is sufficiently low so that the prior does not have much importance whereas in the second case the wrong modeling of the velocity distribution leads to a strong error on the measured  $\Omega_m$ . Fortunately, the scatter distribution presents different types of properties that lead to compatible measurements in Table 6 between the maximum likelihood ( $\sigma_v = +\infty$  to remove the Gaussian prior) and the  $1.5\sigma$  method. For 4k-mock7 and 4k-mock12, the slope estimate is helped by probing velocities with high magnitudes, leading to less possibility of systematic error on the slope.

One is thus led to use a sufficiently deep distance catalogue to ensure the velocity distribution is more or less Gaussian to be able to apply the correction to the likelihood analysis. In this case, one may rely on the value given by  $\mathcal{L}_{\max}$ . If on the contrary, the velocity distribution is highly non-Gaussian, one must use  $\mathcal{L}_{\min}$ . If possible, a visual inspection of the velocity-velocity scatter plot must be conducted to give a check on the amount of statistical biasing.

## CONCLUSION

The Monge-Ampère-Kantorovitch method has been applied with success to reconstruct the velocity field and the density field of simulations (Mohayaee et al. 2006), providing an interesting tool to apply to galaxy catalogues in order to recover the dynamics of our local universe. This method presents the interesting advantage of finding the exact solution of an approximated dynamical problem written in Lagrangian coordinates. The Lagrangian description presents two major advantages. First, it gives a real estimation of peculiar velocities for each galaxies or groups of galaxies, as opposed to a field description which would give an average value at a given spatial position (which is also possible to

build using the Lagrangian description). Second, it permits us to use the Zel’dovich approximation, which gives better peculiar velocity prediction than linear Eulerian theory applied to the same dark matter density field. It means that we expect this method to give better results and more spatially resolved than, *e.g.*, the POTENT method (Bertschinger & Dekel 1989) or velocity field reconstruction through spherical harmonics (Regos & Szalay 1989). Now, most previous analyses of Lagrangian peculiar velocity reconstruction have been run mostly on particle catalogues coming from simulations. However, galaxy catalogues are not as simple, and the main problems are as follows:

- (i) Catalogues mostly provide redshift positions of galaxies and for a few objects their physical distances from us.
- (ii) The luminosity is the only known “dynamical” quantity for most objects in catalogues and so we need extrapolate the M/L relation for known objects to the ones that we do not know.
- (iii) Incompleteness effects have to be taken into account: either because of magnitude limitation or due to extinction of objects by the galactic plane.
- (iv) The MAK reconstruction also needs the Lagrangian domain of the galaxy catalogue.

All these biases and unknown quantities render the reconstruction problem much more difficult than in simulations. We propose here both to test the feasibility of such a reconstruction on galaxy catalogues and the methods to overcome the problems that we have just cited. We have tried to address the following problems:

- Reducing the introduced systematic errors due to unknown bias between mass and luminosity tracers. The dark mass can be either put uniformly into the catalogue or put in the detected haloes (§ 3.1). It appears that there exists an optimum way to distribute the mass, as can be seen in Fig. 5, which gives *unbiased* and *noiseless* reconstructed velocities, even though the exact location of 63% of the mass in the universe remains unknown. In addition to the previous, global, problem, the relative mass distribution between objects in the catalogue is also uncertain as we do not know their true  $M/L$ . The induced systematic errors have been studied in § 3.2 and we show that the naive approach corresponding to using  $M/L = \text{constant}$  inevitably gives a large bias on reconstructed velocities. Even a reasonable guess, for instance the one proposed by Marinoni & Hudson (2002), is still significantly biased. This suggests some more work must be done on the  $M/L$  relation, particularly on the high mass end. However, on the positive side, large random errors on  $M/L$  does not yield any systematic effect and only increases the scatter in the velocity-velocity comparison.

- We proposed a slightly improved way to correct for incompleteness effects in galaxy catalogues and its effect on reconstruction. Though it has given good results, we do not expect this method to be completely bias-free as it presents the same deficiencies as the previous item. However, by enforcing the correction on the *mass* distribution, we managed to preserve the dynamics in the observational data in a better way than would be the case if we had enforced it on the *luminosity* distribution.

- We investigated the eventual systematic errors in redshift reconstructions as proposed previously by Mohayaee &

Tully (2005) and which corresponds to the inverse redshift operator studied by Valentine et al. (2000). It appears that, though the bias is small,  $\Omega_m$  tends to be always underestimated.

- Two solutions to overcome the Lagrangian volume uncertainty for the case of finite volume catalogues have been investigated. The reconstruction method which gives better result seems to be *PaddedDom*. The other alternative, *NaiveDom*, appears to bias the reconstructed velocities, especially in the case of a redshift reconstruction.

- The efficiency of the correction for the zone of avoidance as proposed by Shaya et al. (1995) has been checked (§ 5.1). It appears that the correction is bias free and only introduce a small, but noticeable, additional noise for objects in the direction of the zone of avoidance.

- We checked that the resulting errors of each effect are uncorrelated so they only pile up without producing a strong additional bias. It is fortunate that some observational effects produce complementary biases: incompleteness effect tends to overestimate  $\Omega_m$  whereas redshift distortion underestimates  $\Omega_m$ . The resulting bias is thus not so important.

- We finally tried two estimators to measure  $\Omega_m$  from both reconstructed displacement and distance measurement (§ 6): the  $1.5\sigma$  and the maximum likelihood estimator. However, the first one is not able to work with noisy measured velocities, and the second one is badly affected by large distribution tails in redshift reconstruction. Adding a prior on the distribution of velocities in the catalogue helps to reduce the bias at the cost of having a good measurement of the width of this distribution. A good estimate of  $\Omega_m$  is thus rendered more problematic though we have shown that it should be feasible in principle.

We intend to continue this work in the following directions

- This method can be applied to make a measurement of  $\Omega_m$  in NBG-8k/NBG-3k catalogues and in the upcoming 6dFGS redshift and distance catalogues.

- A better comparison to the acoustic peaks of the CMB can potentially be obtained using the reconstructed displacement field (Eisenstein et al. 2006).

- We can apply MAK reconstruction on SDSS and 2MASS catalogues to obtain the initial Lagrangian positions and velocities of objects in our local universe. This would render the possibility of a re-simulation of our local universe for the first time and check the MAK prediction and correction schemes on real observations.

- We want also to improve the reconstruction itself and propose a new algorithm to include further gravitational effect during orbit reconstructions. This will never give us the internal structure of objects but potentially will give better reconstructed velocities while keeping the power of the MAK reconstruction.

## ACKNOWLEDGEMENTS

We are grateful to S. Prunet, C. Pichon for useful discussions and comments on Maximum Likelihood methods. We would like also to thank D. Weinberg, M. Chodorowski for useful discussions. GL thanks the support and the hospitality of the Institute for Astronomy (University of Hawaii).

This work has been supported by the HORIZON project (<http://www.projet-horizon.fr>).

## REFERENCES

- Peebles P. J. E., 1989, ApJL, 344, L53  
Tully R. B., 2005, ApJ, 618, 214  
Kaiser N., 1987, MNRAS, 227, 1  
Peebles P. J. E., 1980, The large-scale structure of the universe. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.  
Bertsekas D. P., 1979, A Distributed Algorithm for the Assignment Problem. MIT Press, Cambridge, MA  
Strauss M. A., Willick J. A., 1995, Phys. Rept., 261, 271  
Huchra J. P., Geller M. J., Clemens C. M., Tokarz S. P., Michel A., 1992, Bulletin d'Information du Centre de Données Stellaires, 41, 31  
Mohayaee R., Mathis H., Colombi S., Silk J., 2006, MNRAS, 365, 939  
Couchman H. M. P., Thomas P. A., Pearce F. R., 1995, ApJ, 452, 797  
Tully R. B., 1987, ApJ, 321, 280  
Colombi S., Chodorowski M. J., Teyssier R., 2007, MNRAS, 375, 348  
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15  
Mohayaee R., Tully R. B., 2005, ApJL, 635, L113  
Regos E., Szalay A. S., 1989, ApJ, 345, 627  
Efsthathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715  
Bouchet F. R., Colombi S., Hivon E., Juszkiewicz R., 1995, A&A, 296, 575  
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61  
Bertschinger E., Dekel A., 1989, ApJL, 336, L5  
Shaya E. J., Peebles P. J. E., Tully R. B., 1995, ApJ, 454, 15  
Fontanot F., Monaco P., Borgani S., 2003, MNRAS, 341, 692  
Valentine H., Saunders W., Taylor A., 2000, MNRAS, 319, L13  
Zandivarez A., Martínez H. J., Merchán M. E., 2006, ApJ, 650, 137  
Branchini E., Eldar A., Nusser A., 2002, MNRAS, 335, 53  
Croft R. A. C., Gaztanaga E., 1997, MNRAS, 285, 793  
Lahav O., Fisher K. B., Hoffman Y., Scharf C. A., Zaroubi S., 1994, ApJL, 423, L93+  
Marinoni C., Hudson M. J., 2002, ApJ, 569, 101  
Tully R. B., Shaya E. J., Karachentsev I. D., Courtois H., Kocovski D. D., Rizzi L., Peel A., 2007, ArXiv e-prints, 705  
Eisenstein D. J., Seo H.-j., Sirko E., Spergel D., 2006, ArXiv Astrophysics e-prints  
Falco E. E., Kurtz M. J., Geller M. J., Huchra J. P., Peters J., Berlind P., Mink D. J., Tokarz S. P., Elwell B., 1999, PASP, 111, 438  
Phelps S. D., Desjacques V., Nusser A., Shaya E. J., 2006, MNRAS, 370, 1361  
Tully R. B., Pierce M. J., 2000, ApJ, 533, 744



## Observational biases in Lagrangian reconstructions of cosmic velocity fields 27

da Costa L. N., Pellegrini P. S., Sargent W. L. W., Tonry J., Davis M., Meiksin A., Latham D. W., Menzies J. W., Coulson I. A., 1988, *ApJ*, 327, 544  
 Brenier Y., Frisch U., Hénon M., Loeper G., Matarrese S., Mohayaee R., Sobolevskii A., 2003, *MNRAS*, 346, 501  
 Pizagno J., Prada F., Weinberg D. H., Rix H.-W., Pogge R. W., Grebel E. K., Harbeck D., Blanton M., Brinkmann J., Gunn J. E., 2006, *ArXiv Astrophysics e-prints*  
 Erdoğan P., Lahav O., Huchra J. P., Colless M., Cutri R. M., Falco E., George T., Jarrett T., Jones D. H., Macri L. M., Mader J., Martinbeau N., Pahre M. A., Parker Q. A., Rassat A., Saunders W., 2006, *MNRAS*, 373, 45  
 Spergel D. N., Bean R., Dore O., Nolte M. R., Bennett C. L., Hinshaw G., Jarosik N., Komatsu E., Page L., Peiris H. V., Verde L., Barnes C., Halpern M., Hill R. S., Kogut A., Limon M., Meyer S. S., Odegard N., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2006, *ArXiv Astrophysics e-prints*  
 Bennett C. L., Halpern M., Hinshaw G., Jarosik N., Kogut A., Limon M., Meyer S. S., Page L., Spergel D. N., Tucker G. S., Wollack E., Wright E. L., Barnes C., Greason M. R., Hill R. S., Komatsu E., Nolte M. R., Odegard N., Peiris H. V., Verde L., Weiland J. L., 2003, *ApJS*, 148, 1  
 Tegmark M. et al., 2006, *Phys. Rev. D*, 74, 123507

## APPENDIX A: CONSTRUCTION OF A MAK MESH

MAK reconstructions requires a sampling of the matter distribution with “particles” of equal mass corresponding to nodes of an homogeneous mesh. When considering the simulation, one uses a full periodic cubic mesh. However, in real galaxy catalogues, the relevant lagrangian volume is a non-periodic compact subset inscribed in a larger rectangular mesh. In that case, the assignment is performed only for “particles” belonging to this initial volume. Note that the determination of this initial volume is by itself a great challenge and a poor guess can have dramatic consequences.

Given a number of “galaxies”, or tracers, for which the individual masses  $M_i$  are known and a choice of the mass resolution of the MAK grid,  $m_R$ , the problem is now to determine how many “particles” have to be assigned to tracer  $i$ . This number should be  $n_i = M_i/m_R$  which is rarely an integer. To address this issue, we construct an integer function  $\tilde{n}_i$  such that the quantity

$$\chi^2 \equiv \sum_i \left( \frac{\tilde{n}_i m_R - M_i}{M_i} \right)^2 \quad (\text{A1})$$

is minimized given the constrain

$$\sum_i \tilde{n}_i = N_{\text{MAK}}, \quad (\text{A2})$$

where  $N$  is the total number of nodes on the MAK grid, such that  $N_{\text{MAK}} \times m_R$  is as close as possible to the total mass,  $\sum M_i$ . The minimization of  $\chi^2$  is performed iteratively until convergence. Note that the solution of such a minimization is, in general, not unique due to the possible permutations between objects of the same mass. Due to this degeneracy, it is needed to shuffle randomly the tracers prior to the minimization in order to avoid possible systematic effects.

Note finally that one must make sure that there is at least a few particles per tracer,  $n_i \geq \alpha$  with  $\alpha > 1$ . This brings constraints on  $m_R$  and therefore on the size of the MAK mesh. Unfortunately, it is not always possible to have  $\alpha > 1$  due to the prohibitive CPU cost it would imply for the MAK reconstruction in the present paper. To address this problem, we separate the catalogue into groups of galaxies and field galaxies. For the groups, the  $\chi^2$  minimization is performed as explained above, with a possible loss of the lightest ones since  $n_i$  can still be smaller than unity. For the field galaxies, we use a simpler procedure as follows. Given the mass  $M_i$  of a galaxy  $i$ , a MAK tracer is randomly assigned to it with occurrence probability  $M_i/m_R$ .

## APPENDIX B: TOOLS FOR ERROR ANALYSIS

To check the accuracy of the reconstructions, we compute the moment of the joint probability distribution of the reconstructed velocities  $\mathbf{v}_{\text{rec},i}$  of object  $i$  and the simulated velocities of those objects  $\mathbf{v}_{\text{sim},i}$ . We write  $\langle A \rangle$  the average of the quantity  $A$

$$\langle A \rangle = \frac{1}{N} \sum_{i=1}^N A_i, \quad (\text{B1})$$

We define three second moments (after subtraction of the average):

$$\sigma_r^2 = \langle v_{\text{rec}}^2 \rangle, \quad \sigma_s^2 = \langle v_{\text{sim}}^2 \rangle, \quad \sigma_{r,s} = \langle v_{\text{rec}} v_{\text{sim}} \rangle \quad (\text{B2})$$

From these moments we can build the correlation coefficient:

$$r = \frac{\sigma_{r,s}}{\sigma_r \sigma_s} \quad (\text{B3})$$

and the ratio between the width of the reconstructed field PDF (density or velocity) and the width of simulated – mock – field PDF

$$s = \frac{\sigma_r}{\sigma_s}. \quad (\text{B4})$$

For these two quantities the optimum value is 1. Alternatively two other “slope” estimator of the reconstructed velocities versus the simulated ones can be built from the above momenta

$$s_{\text{min}} = \frac{\sigma_{r,s}}{\sigma_s^2} = sr \quad \text{and} \quad s_{\text{max}} = \frac{\sigma_r^2}{\sigma_{r,s}} = s/r. \quad (\text{B5})$$

These two slopes are interesting when one makes an estimation of  $\Omega_m$  through  $s$  and needs an evaluation of the uncertainty. The two extra slopes determined using this way should, ideally, be equal to  $s$  but due to the lack of perfect correlation ( $r < 1$ ), they are actually different from it in realistic cases. In fact, we have  $s_{\text{min}} < s_{\text{med}} < s_{\text{max}}$ .

Please note that we can define the relative dispersion

$$\sigma^2 = \frac{\langle (v_{\text{rec}} - v_{\text{sim}})^2 \rangle}{\sigma_s^2} = 1 + s^2 - 2sr, \quad (\text{B6})$$

which is a measure of the noise-to-signal ratio: high  $\sigma$  corresponds to low signal. Ideally, one wants  $\sigma = 0$ .

## APPENDIX C: SIMULATING MAGNITUDE-LIMITED CATALOGUES

Having only a halo catalogue, we must generate a “galaxy catalogue” including incompleteness effects. The main diffi-

culty in that construction is that the distribution of galaxies in the universe is a non-trivial, non-linear functional of the total matter density field. For instance, bright galaxies tend to concentrate in massive structures (Zandivarez et al. 2006). It means that, though most of the field galaxies are missed, the major groups can still be easily seen due to the bright galaxies they contain. Thus the galaxy distribution should mostly trace large haloes at large distances, potentially introducing a bias in the reconstructed peculiar velocities if incompleteness corrections are performed unwisely. In what follows, we generate mock galaxy catalogues like NBG-8k/3k. To take properly into account the effects discussed above, we separate groups of galaxies from field galaxies. Groups are populated with galaxies following the universal Schechter form for simplicity, but with a different normalization to account for their non-trivial  $M/L$ .

Statistically, NBG-8k/3k catalogues are composed of galaxies measured in the B band and distributed according to the Schechter form

$$n(L) dL \simeq n_0 L^{-1} e^{-L/L_*} dL, \quad (C1)$$

with  $L_* \simeq 5.7 \times 10^{10} L_\odot$  and  $n_0 \simeq 0.03 h^3 \text{Mpc}^{-3}$ . Moreover, the NBG-8k catalogue is complete above  $3 \times 10^9 - 4 \times 10^9 L_\odot$  inside a sphere of radius  $d_{\text{comp}} = 12 h^{-1} \text{Mpc}$ . As the mean “galaxy” (particle) density in the simulation is  $n_{\text{sim}} = 0.26 h^3 \text{Mpc}^{-3}$  and about  $n_{\text{cat}} = 0.08 h^3 \text{Mpc}^{-3} \simeq 0.30 n_{\text{sim}}$  in NBG-8k, we must dilute the simulation to get a mock catalogue similar to NBG-8k. The luminosity  $L_G$  of a detected galaxy at a distance  $d$  from the observer must satisfy the constraint

$$L_G > 4\pi l_{\text{cut}} d^2 \quad (C2)$$

with  $l_{\text{cut}}$  the minimum flux detectable by the observer. The fraction of galaxies detected at the distance  $d$  in the galaxy mock catalogue is thus

$$f_{\text{field}}(d) = \begin{cases} 0.30 & \text{if } d < d_{\text{comp}} \\ \frac{\int_{4\pi l_{\text{cut}} d^2}^{\infty} n(L) dL}{\int_{L_{\text{min}}}^{\infty} n(L) dL} & \text{otherwise} \end{cases} \quad (C3)$$

with  $l_{\text{cut}}$  the minimum flux detectable by the observer. The fraction is saturated at 0.30 to follow the dilution constraint expressed above. We enforce the continuity of  $f_{\text{field}}(d)$  by choosing  $L_{\text{min}}$  such that  $f_{\text{field}}(d_{\text{comp}}) = 0.30$ .

The mock galaxy and group of galaxies catalogue is now built:

**I.** We take a halo  $A$  from *FullMock* and assume it is a group of galaxies. We thus deduce the intrinsic luminosity  $L_A$  from the mass  $M_A$  of this object using Eq. (5).

**II.** The observed luminosity  $L'_A$  of  $A$  is computed assuming that its galaxy population follows (C1) but with a different normalization to achieve the intrinsic luminosity  $L_A$ . If  $d_A$  is the distance between the observer and the halo  $A$ , then the galaxies detected in this halo verify (C2) for  $d = d_A$ . The total observable luminosity for  $A$  is thus

$$L'_A = L_A f_L(d_A) \quad (C4)$$

with, assuming  $L_{\text{min}} \ll L_*$ ,

$$f_L(d) = \begin{cases} \sim 1 & d < d_{\text{comp}} \\ \frac{\int_{4\pi d^2 l_{\text{cut}}}^{\infty} L n(L) dL}{\int_{L_{\text{min}}}^{\infty} L n(L) dL} \simeq e^{-\frac{4\pi d^2 l_{\text{cut}}}{L_*}} & d \geq d_{\text{comp}} \end{cases} \quad (C5)$$

**III.** If  $L'_A < 4\pi d_A^2 l_{\text{cut}}$  then  $A$  is removed from the catalogue, otherwise it is kept.

**IV.** This gives us the group component of our magnitude-limited catalogue.

**V.** The case of the “field galaxies” is treated separately. Galaxies are identified with dark matter particles and their luminosity is assigned following (C1). More specifically, we choose a shell  $S_d$  put at a distance  $d$  from the observer. The probability of keeping a “galaxy”  $G$  in  $S_d$  is given by (C3). Inside the shell  $S_d$ , the selected “galaxies” share now a luminosity

$$L_{\text{f}}(d) = \int_{4\pi d^2 l_{\text{cut}}}^{\infty} L n(L) dL \quad (C6)$$

which is distributed evenly among them. Strictly speaking, such a repartition should be performed randomly according to (C1). That would add a small additional noise on the reconstructed velocities. This noise should be of insignificant consequence as supported by the discussion of the *TS-T* case in § 3.2.

We have now a realistic mock catalogue and we try to account for its incompleteness as we would for NBG-8k:

**A.** The missing luminosity in groups is corrected. In order to do this, we compute, in a thin shell  $S_d$  at some distance  $d$ , the ratio between the expected total luminosity and the observed luminosity

$$b(d) = \frac{\int_0^{\infty} L n(L) dL}{\int_{4\pi d^2 l_{\text{cut}}}^{\infty} L n(L) dL} = e^{4\pi d^2 l_{\text{cut}} / L_*}. \quad (C7)$$

The intrinsic luminosity  $L_A$  of a group  $A$  in  $S_d$  is recovered with

$$L_A = L_{\text{obs},A} b(d). \quad (C8)$$

The mass  $M_A$  of  $A$  can then be obtained using the *non-linear* relation (5).

**B.** The remaining missing mass in  $S_d$  can be written

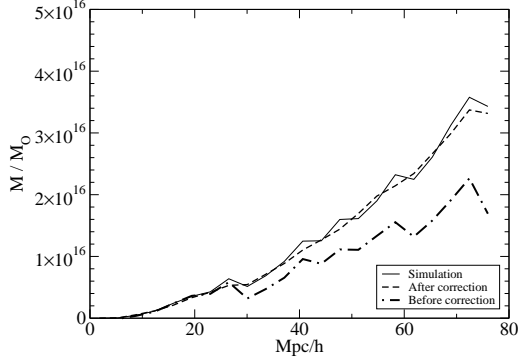
$$M_{\text{missed},d} = \Upsilon b(d) (L_{\text{field,obs},d} + L_{\text{group,obs},d}) - M_{\text{field,obs},d} - M_{\text{group,obs},d}, \quad (C9)$$

with  $\Upsilon = 93 \frac{M_\odot}{L_\odot}$  the average  $M/L$ ,<sup>18</sup>  $L_{\text{group,obs},d}$  the observed luminosity of groups,  $M_{\text{group,obs},d}$  the masses of groups obtained after the above correction,  $L_{\text{field,obs},d}$  the luminosity of field galaxies. The quantity  $M_{\text{missed},d}$  comes from both missing galaxies and missing group of galaxies. If  $M_{\text{missed},d} > 0$  and without any further information, the missing mass may either be assigned evenly to field galaxies of  $S_d$  (our choice, as usually performed in the literature), or distributed uniformly in  $S_d$  using new random tracers. If  $M_{\text{missed},d} \leq 0$ , the mass distribution in  $S_d$  is untouched.

This procedure is certainly not free from biases. For instance, the contrasts between shells are partly smoothed out, as illustrated by Fig. C1. This is equivalent to reducing the overall magnitude of fluctuations in the density field. As a result, a small bias towards larger  $\Omega_m$  might occur, as in the lower right panel of Fig. 5 of § 3.1. On the opposite, if

<sup>18</sup> Note that a prior assumption on the value of  $\Omega_m$  is obviously needed to estimate  $\Upsilon$ .





**Figure C1.** *Magnitude limitation/Filling missing mass* – This plot gives the measured amount of mass in a thin shell at different distances from the observer. The solid line gives the original mass distribution in the simulation, the dot-dashed line the mass distribution after mimicking incompleteness and the dashed line the recovered mass distribution after correction for incompleteness as described in Appendix C.

the missing mass is assigned to detected background galaxies, the estimated  $\Omega_m$  is expected to underestimate the true value as discussed in § 3.1.

C. Note that the mass of the “field galaxies” is not the mass of a single particle anymore. Procedure explained in Appendix A is facilitated as follows, for the sake of algorithmic simplicity. With  $v \in (0; 1]$  a uniform random variable, a galaxy  $G$  of mass  $m_G$  is splitted into  $n_G$  subcomponents of mass  $m_{\text{particle}}$  such that:

$$n_G = \begin{cases} r_G & \text{if } \left( \frac{m_G}{m_{\text{particle}}} - r_G \right) < v \\ r_G + 1 & \text{otherwise} \end{cases}, \quad (\text{C10})$$

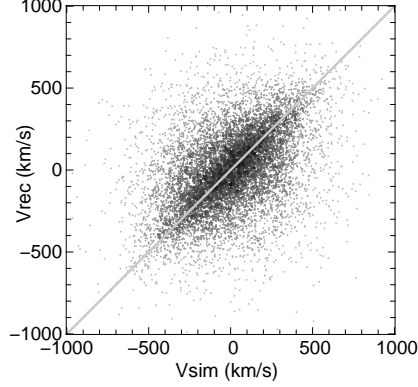
with  $r_G = \left\lfloor \frac{m_G}{m_{\text{particle}}} \right\rfloor$ ,  $\lfloor x \rfloor$  being the integer part of  $x$ . Each of the subcomponent is now considered as a “field galaxy” in the procedure explained in Appendix A.

#### APPENDIX D: STATISTICAL BIAS IN THE SLOPE ESTIMATION

The two methods that we used for slope estimation are known to be biased. A more precise treatment of this bias is beyond the scope of this paper. However we propose here to check the order of magnitude of the systematic effect of the statistical analysis itself. To achieve this we produced a set of randomly generated “velocities”  $v$  and their “reconstructed velocities”  $v_R$  counterpart. The probability for a point  $(v, v_R)$  to have a velocity  $v$  is given by

$$P_v(v) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-v^2/(2\sigma_v^2)}, \quad (\text{D1})$$

with  $\sigma_v = 300 \text{ km s}^{-1}$  typically. The probability for it to have a reconstructed velocity  $v_R$  is given by the *same* probability law. We now compute the error  $e$  between  $v_R$  and  $v$ , which



**Figure D1.** *Statistical bias* – Scatter plot of 10,000 randomly generated points following the approximated probability laws found between reconstructed velocities and simulated velocities.

must be distributed according to the Lorentzian form

$$P_{DE}(e) = \frac{1}{\pi B} \frac{1}{1 + \left(\frac{e}{B}\right)^2}, \quad (\text{D2})$$

with  $B = 86 \text{ km s}^{-1}$ . The error  $e$  is related to  $v$  and  $v_R$  by

$$e = \alpha_* v - \beta_* v_R. \quad (\text{D3})$$

For the rest of the appendix we take  $\alpha_* = \beta_* = 1/\sqrt{2}$ . The probability of keeping a point  $(v, v_R)$  with an error  $e$  is given by, integrating  $P_{DE}(e')$  between  $-e$  and  $+e$ ,

$$P_{\text{keep}}(e) = \frac{2}{\pi} \tan^{-1} \left( \frac{e}{B} \right) \quad (\text{D4})$$

We represented in Fig. D1 a scatter plot of 10,000 points generated using this procedure. As one can see, it does look like a real scatter plot of a redshift reconstruction.

Conducting a  $1.5\sigma$  analysis on this set of points, we find a slope  $\beta_*/\alpha_* = 1.0 \pm 0.20$ . In our case, this would give  $\Omega_m = 0.30 \pm 0.10$ . Estimating the slope using the maximum likelihood approach gives, with  $\sigma_v = +\infty$ ,  $\beta_*/\alpha_* = 0.81 \pm 0.01$  ( $\Omega_m = 0.20 \pm 0.02$ ) and with  $\sigma_v = 300 \text{ km s}^{-1}$ ,  $\beta_*/\alpha_* = 1.074 \pm 0.012$  ( $\Omega_m = 0.34 \pm 0.02$ ). Putting  $B = 40 \text{ km s}^{-1}$ , both for generated data and likelihood function, as for real space reconstructions, reduces the error and gives  $\Omega_m = 0.31 \pm 0.02$ , thus highlighting the importance of the reconstruction noise for a good estimation of  $\Omega_m$ .

Consequently, though one must rely on the likelihood analysis, it may be strongly biased by the structure of reconstruction errors mixed with the non-uniform distribution of observables. We tried to make a good approximate model of the errors, though it seems to quite depends on the value of  $\sigma_v$ . Whenever possible, of course, one must crosscheck the result of the likelihood by a visual inspection of the scatter plot.

## **Annexe I Article soumis**

DRAFT VERSION MAY 20, 2008  
 Preprint typeset using L<sup>A</sup>T<sub>E</sub>X style emulatej v. 12/14/05

## 2MASS REDSHIFT SURVEY: THE ORIGIN OF THE CMB DIPOLE

GUILHEM LAVAUX<sup>1</sup>, R. BRENT TULLY<sup>2</sup>, ROYA MOHAYAE<sup>1</sup>, STEPHANE COLOMBI<sup>1</sup> (THE 2MRS COLLABORATION)

*Draft version May 20, 2008*

### ABSTRACT

We present the peculiar velocity field for the 2MASS Redshift Survey (2MRS) catalog using an orbit-reconstruction algorithm. The reconstructed velocities of individual objects in 2MRS are well-correlated with the peculiar velocities obtained from high-precision observed distances within 3000 km s<sup>-1</sup>. The reconstructed motion of the Local Group in the rest frame established by distances within 3000 km s<sup>-1</sup> agrees with the observed motion and is generated by fluctuations within this volume, in agreement with observation. Having tested our method against observed distances, we reconstruct the velocity field of 2MRS in successively larger radii, to study the problem of convergence towards the CMB dipole. We find that less than half of the amplitude of the CMB dipole is generated within a volume enclosing the Hydra-Centaurus-Norma supercluster at around 40h<sup>-1</sup> Mpc. Although most of the amplitude of the CMB dipole seems to be recovered by 120h<sup>-1</sup> Mpc, we observe no convergence up to this scale. Due to dominant superclusters such as Horologium-Recticulum in the southern hemisphere at scales above 120h<sup>-1</sup> Mpc, one might need go well beyond the Shapley supercluster to fully recover the dipole vector.

*Subject headings:* dark matter — methods:analytical, numerical and observation

### 1. INTRODUCTION

The Cosmic Microwave Background (CMB) dipole indicates that our Local Group (LG) moves with a velocity of 627±22 km s<sup>-1</sup> towards the direction  $l = 276^\circ$ ,  $b = 30^\circ$  in galactic coordinates. It is highly unlikely that the Hydra-Centaurus-Norma supercluster (HCN) at around 40h<sup>-1</sup> Mpc could be responsible for the dipole and it is now widely believed that one has to go as far as the Shapley concentration at about 150h<sup>-1</sup> Mpc in order to fully recover the dipole motion. However, due mainly to sparseness of data at very large distances, contradictory results have been reported and so far there is no consensus on the depth of the convergence for the CMB dipole (Lauer & Postman 1994; Giovanelli et al. 1996; Kocevski & Ebeling 2006).

The dipole motion, generated by large-scale inhomogeneities, is obtained from the peculiar velocity field either directly by observing these velocities through distance measurements or indirectly by extracting these velocities from redshift surveys. The catalog of distances become increasingly sparse and inaccurate as we go farther away. Since only the radial distances are observed, a  $\chi^2$  minimization is usually used to infer the bulk velocity which can be inaccurate due to projection effects (Regos & Szalay 1989; Colless 1995).

The other common approach to recovery of the CMB dipole has been to use linear theory to reconstruct the velocity field from the redshift surveys (see *e.g.* Erdoğdu et al. 2006a, 2006b; Kocevski & Ebeling 2006). Methods based on linear theory can suffer from inadequacy in dealing with large fluctuations such as the HCN, redshift space distortion and also the fact that the reconstructed velocity field is never tested against any observations.

Here, we apply the Lagrangian method, MAK, of peculiar-velocity reconstruction to the 2MRS catalog and produce a 3-dimensional map of the velocity field. The

method has been adapted to work directly with redshifts and allows us to go well beyond the linear theory into the non-linear regime. The method has been tested previously against simulations and mock catalogs and has been shown to reconstruct reliable peculiar velocities on scales above 4-5 Mpc (see Lavaux et al. 2008 and references therein). In addition, using this method we can directly generate the 3-components of the peculiar velocities and hence overcome projection effects.

The reconstructed velocities of the 2MRS galaxies are tested against observed velocities obtained from high-precision distance measurements within a 3000 km s<sup>-1</sup>(3k) radius (Tully et al. 2008). The two velocities are well-correlated with small dispersion. The second test consists of comparing the reconstructed velocity of Local Group in the rest frame of the 3k sample to the observed velocity. Our results show that the LG motion in the rest frame of 3k is mostly generated within this volume, as indicated by the observations (Tully et al. 2008).

Having tested our method, we study the origin of the CMB dipole. We reconstruct the velocity field of 2MRS in successively larger radii in order to determine whether there is any trend of convergence towards the CMB dipole. Contrary to previous results on 2MRS (Erdoğdu et al. 2006b), we show that the depth of the convergence lies beyond 120h<sup>-1</sup> Mpc. Due to severe incompleteness of the 2MRS catalog beyond 120h<sup>-1</sup> Mpc, we can only put a lower-limit on the value of the convergence depth. Our method allows us to determine the rate at which the dipole is approached: less than 50% of the dipole amplitude is achieved at around the HCN distance and at around 120h<sup>-1</sup> Mpc and we recover about 87% of the amplitude of the CMB dipole but with no evidence for convergence in direction.

<sup>1</sup> Institut d'Astrophysique de Paris, 98bis bd Arago, France

<sup>2</sup> Institute for Astronomy, Univ. of Hawaii, Honolulu, USA

2. THE COMBINED CATALOG: 2MASS REDSHIFT CATALOG (2MRS) PLUS 3K CATALOG OF DISTANCES

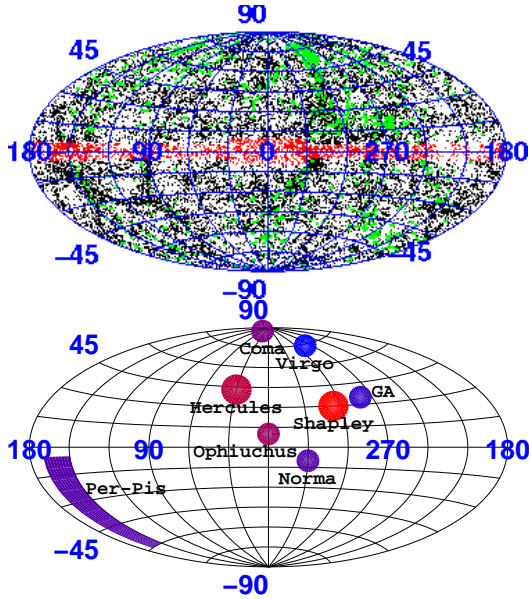


FIG. 1.— Combined 2MRS redshift catalog and 3k catalog of distances: Top panel: Large solid green circles are the galaxies for which a distance measure is available. The zone of avoidance has also been filled and is shown by the red belt running across the catalog. Bottom panel: angular positions of a few dominant superclusters in 2MRS catalog with more distant objects shown by redder circles.

The 2MRS is the most uniformly sampled all-sky redshift survey to date and its selection in the near infra-red reduces the impact of the zone of avoidance. The catalog is based on galaxy selection in the infra-red from the Two Micron All-Sky Survey (2MASS). The  $K=11.25$  mag 2MRS is now complete and contains about 23,200 galaxies. 2MRS provides good coverage to a mean redshift of  $z=0.02$  and becomes extremely sparse afterwards up to  $200h^{-1}$  Mpc.

We fill the zone of avoidance, which mainly extends over  $|b| < 5 - 10^\circ$ , by mirroring galaxies from the unmasked region (see Shaya et al. 1995 and Lavaux et al. 2008 for details). The kinematic fingers of god (Jackson 1972), associated with clusters, are compressed using the algorithm of Huchra & Geller (1982). We used a fiducial velocity  $V_F = 1000 \text{ km s}^{-1}$ ,  $V_0 = 350 \text{ km s}^{-1}$ ,  $D_0 = 0.41 h^{-1} \text{ Mpc}$ , which corresponds to detecting galaxy count overdensities of  $\delta n/\bar{n} \simeq 80$  and provide robust statistical properties of groups as explained by Ramella et al. (1997). We assume a constant  $M/L$  ratio as it has been shown that the masses of big groups of galaxies only moderately depend on luminosities (Lin et al. 2004). Incompleteness is taken into account by using the Schechter luminosity function given in Crook et al. (2007) and Huchra et al. (2005) for 2MRS. The specific values of the parameters are:  $\Phi_* = 1.06 \cdot 10^{-2} h^3 \text{ Mpc}^{-1}$ ,  $M_* = -24.2 + 5 \log_{10} h_{73}$ . We have verified that, as far as galaxies of the 2MRS are concerned, there does not seem to be any systematic effect in the relation between the  $K_{20}$  magnitude and the intrinsic luminosity.

The second observational component is an extended catalog of galaxy distances. Information from four techniques has been integrated: the Cepheid variable (Freedman et al. 2001), Tip of the Red Giant Branch (Karachentsev et al. 2003; Lee et al. 1993), Surface Brightness Fluctuation (Tonry & Schneider 1988; Tonry et al. 2001), and Luminosity–Linewidth (Tully & Fisher 1977; Tully & Pierce 2000) methods. In all, there are 1791 galaxies with distance measures within  $3,000 \text{ km s}^{-1}$  (whence the name “3k catalog”); over 600 of these are derived by at least one of the first three ‘high quality’ techniques. The 3k catalog of distances has been described in detail by Tully et al. (2008).

The combined 2MRS catalog and 3k distance catalog are shown in Figure 1. In the present study, distances are averaged over groups because our method cannot meaningfully recover the velocities on sub-group scales. The present mixed catalog is composed of 24,819 galaxies, among which 1,126 has been assigned a distance. Distances are assigned to 109 groups of galaxies out of 695 groups which leaves 617 galaxies with measured distances but which have not been grouped.

### 3. TECHNIQUE: MAK RECONSTRUCTION OF PECULIAR VELOCITY FIELD

The technique that will be implemented here is the Monge–Ampère–Kantorovich (MAK) reconstruction. This method provides a recipe for reconstructing galaxy orbits that is unique to the degree that orbits can be described as following straight lines under suitable coordinate transformations (Lavaux et al. 2008, and references therein). In essence, MAK searches for the minimum of the discretized action

$$S = \sum_{i=1}^N \left( \mathbf{q}_{j(i)} - \mathbf{x}_i \right)^2. \quad (1)$$

which assigns initial positions  $\mathbf{q}_j$  to final positions  $\mathbf{x}_i$ . The action (1) can be modified easily to account for redshift distortions (Mohayaee & Tully 2005; Lavaux et al. 2008). The peculiar velocities and also the relationship between redshift and real space displacements of the objects (galaxies or groups) is estimated using the Zel’dovich approximation (Zel’dovich 1970):  $\mathbf{v} = f\mathbf{H}(\mathbf{x} - \mathbf{q})$  where  $\mathbf{v}$  is the peculiar velocity vector and  $f \sim \Omega_m^{0.6}$ , and  $\mathbf{x}$  is the reconstructed real position computed by MAK. Consequently, a specific model defines positions that can be tested against observed positions, as shall be pursued here. The method has already been tested successfully on N-body simulations (Mohayaee et al. 2006), mock catalogs (Lavaux et al. 2008) and a real catalog (Mohayaee & Tully 2005).

### 4. RESULTS: I. 2MRS VELOCITY FIELD AND COMPARISON WITH MEASURED DISTANCES

To reconstruct the velocity field using MAK, as discussed in Section 3, we fix  $\Omega_m = 0.258$  to the WMAP5 results and set the bias to 1 using 2dF and SDSS/WMAP results (Dunkley et al. 2008; Tegmark et al. 2004; Cole et al. 2005). The velocity of each galaxy in 2MRS is reconstructed using these parameters. In a parallel approach, we leave  $\Omega_m$  free and then constrain its value by maximising the correlation between the reconstructed and observed peculiar velocities. We obtain an error estimated at  $\sim 5\%$  on reconstructed peculiar velocities due

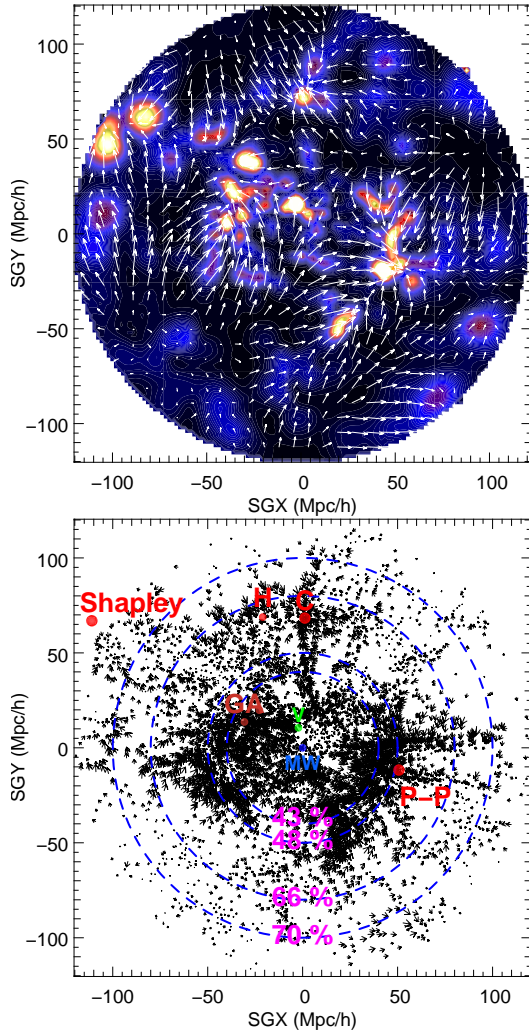


FIG. 2.— *2MRS peculiar velocity field*: Top panel shows a thin slice of the 2MRS peculiar velocity field adaptively smooth on a grid with  $128^3$  sites sampling a cube with an edge of  $240 h^{-1}$  Mpc. The velocity field is then subsampled 4 times before being shown. The underlying density field has been computed by putting objects at their redshift position. Bottom panel: Peculiar velocities of individual 2MRS galaxies in a  $40 h^{-1}$  Mpc slice, centered on the supergalactic plane  $SGZ=0$  km s $^{-1}$ . 2MRS becomes severely incomplete after  $120 h^{-1}$  Mpc.

to the assumed values of cosmological parameters. The 2MRS velocities are reconstructed using a uniform grid of size  $130^3$  sampling a cubic volume of  $260^3 h^{-3} \text{Mpc}^3$  as shown in Fig. 2. The motion of the LG is obtained using an interpolation based on the peculiar velocities of the objects that lie within  $4\text{--}5 h^{-1}$  Mpc from us. We have checked that increasing the reconstruction resolution does not significantly change the reconstructed velocities.

The reconstructed velocities of objects lying only

within the  $3,000 \text{ km s}^{-1}$  radius can be compared to the measured distances given by the 3k distance catalog. The measured distances give a velocity of the LG with respect to the 3k volume of  $V_{\text{LG}/3\text{k}} = 302 \pm 22 \text{ km s}^{-1}$ ,  $l = 241 \pm 7$ ,  $b = 37 \pm 6$ . The observation also indicates that most of this velocity comes from the push from the Local Void and the gravitational pull of the Virgo cluster (Tully et al. 2008). The LG velocity with respect to the 3k volume ( $V_{\text{LG}/3\text{k}}$ ) is obtained using our reconstructed 2MRS velocities. The reconstructed  $V_{\text{LG}/3\text{k}}$  is compared with the observed value in Fig. 3. The reconstruction indeed shows that the  $V_{\text{LG}/3\text{k}}$  motion is mainly generated within the 3k volume and is only marginally due to HCN at  $40\text{--}50 h^{-1}$  Mpc. Reconstructed velocities are well-correlated with those obtained from measured distances, with a correlation coefficient of  $r = 0.69$  (for optimal correlation  $r = 1$ ) and a typical relative dispersion of  $\sigma = 128 \text{ km s}^{-1}$  (see Lavaux et al. 2008 for a detailed definition of these parameters). Having obtained and tested our reconstructed velocities within  $3,000 \text{ km s}^{-1}$ , we go to larger scales and study the origin of the LG motion with respect to CMB,  $V_{\text{LG}/\text{CMB}}$ , in the next section.

#### 5. RESULTS: II. 2MRS VELOCITY FIELD AND THE ORIGIN OF THE CMB DIPOLE

The CMB dipole motion, is obtained by using the reconstructed 3d velocity field. The velocities are reconstructed within increasing radii centred on the LG. A plot of the resulting dipole as a function of the reconstruction radius is made in Fig. 4. No convergence is achieved below  $120 h^{-1}$  Mpc. The 2MRS sample becomes highly incomplete beyond this scale and further conclusions cannot be made before a more complete sample becomes available. Whether the Shapley supercluster yields the convergence to the CMB dipole also remains questionable, as farther away in the southern hemisphere, dominant structures such as the Horologium-Reticulum supercluster could change the direction of the dipole. Thus, it is possible that the depth of the convergence towards CMB dipole lies well-beyond the Shapley concentration itself.

We also plot the increase in the bulk flow relative to the CMB dipole in the bottom panel of Fig. 2. The figure demonstrates that less than half of the amplitude of the dipole is generated within the volume enclosing HCN. To reach convergence a significant contribution to the dipole has to be made from the Shapley supercluster and beyond. This agrees with an analysis of the X-ray data (Kocevski & Ebeling 2006). We also fail to observe convergence by the time we reach  $120 h^{-1}$  Mpc. Indeed, although the amplitude of the reconstructed dipole seems to approach that of CMB, its direction remains well away from it. Using observational error bars, we estimate that the observational error on the direction of the LG velocity should not exceed  $\sim 6^\circ$  at 95% confidence. In addition, there is an intrinsic  $22^\circ$  error (95%) in the direction of the reconstructed velocity because of the modeling errors. So we do not expect the two vectors to be more than  $\sim 28^\circ$  apart (though here they lie at  $\sim 50^\circ$  from each other).

#### 6. CONCLUSION

We have evaluated the 3-dimensional 2MRS peculiar velocity field using a Lagrangian method (MAK) which

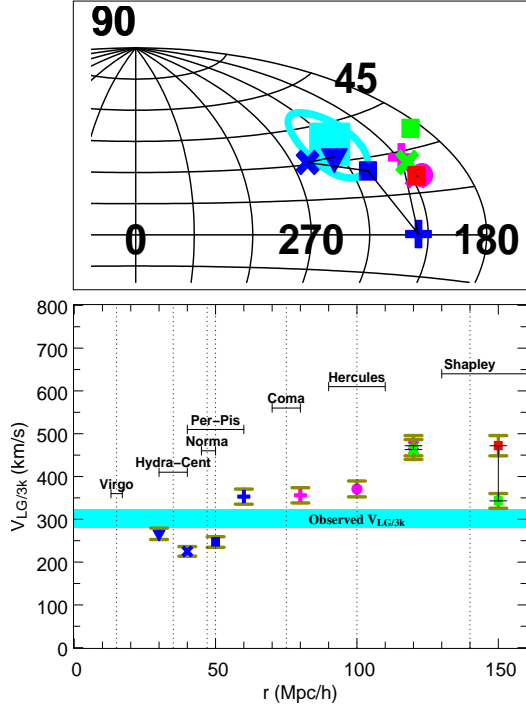


FIG. 3.— The origin of LG motion in the rest frame of galaxies in the  $3000 \text{ km s}^{-1} (3k)$  volume,  $V_{LG/3k}$ : The observed amplitude and direction of the  $V_{LG/3k}$  are shown by the horizontal cyan band in the lower panel and the solid cyan large square in the top panel, respectively. The observational error on the direction is shown on the top panel by the thick cyan ellipse. Each reconstructed point is affected by a  $22^\circ$  error. The reconstructions indicate that most of the motion is generated by the structures within this volume, in agreement with the observations. It is seen that most of the motion is generated within  $3,000 \text{ km s}^{-1}$ . The jump at around  $120 h^{-1} \text{ Mpc}$  is most probably caused by the severe incompleteness of 2MRS at this distance. The 5% error bar on each of the amplitude measurements comes from uncertainty in  $\Omega_m$ . Incompleteness results in uncertainties illustrated by the black error bars on data points beyond  $120 h^{-1} \text{ Mpc}$  in the bottom panel.

works well into the nonlinear regime, on scales above  $4.5 h^{-1} \text{ Mpc}$ . The method has been adapted to work directly with redshifts. The reconstructed velocities are well-correlated with the distances observed within a radius of  $3k$ . The reconstructed velocity of the Local Group in the rest frame of  $3k$  also agrees well with the observed value. We have then studied the origin of the Local Group motion in the CMB rest frame by using our 3d reconstructed velocities. We have shown that less than half of the CMB dipole could be generated within a radius enclosing the Hydra-Centaurus-Norma supercluster. We have demonstrated how the trend of the

convergence varies up to  $120 h^{-1} \text{ Mpc}$  and have shown that convergence in position to the CMB dipole is not reached below this distance. Forthcoming deeper and more complete redshift surveys especially those in X-ray, could soon establish whether one has to go well beyond  $120 h^{-1} \text{ Mpc}$  to recover the CMB dipole or whether the Shapley concentration at around  $150 h^{-1} \text{ Mpc}$  is sufficient to finally reach the convergence. It remains to be

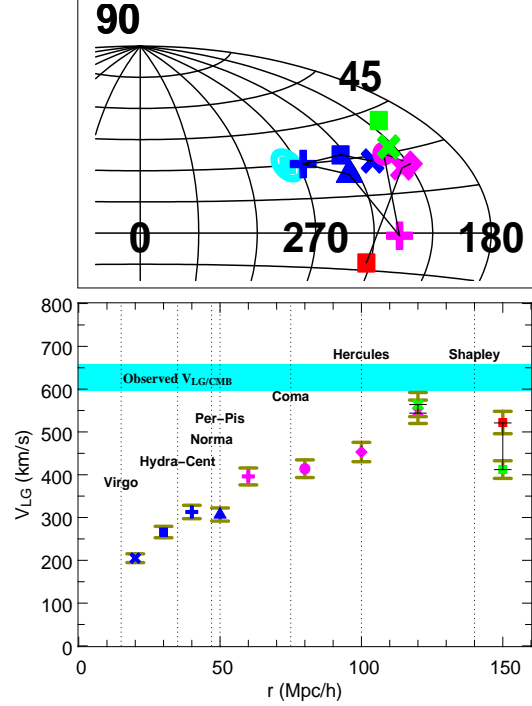


FIG. 4.— LG velocity in progressively larger rest frames,  $V_{LG}$ : The observed amplitude and direction of the CMB dipole motion are shown by the horizontal cyan band in the lower panel and the solid cyan large square in the top panel, respectively. The lower panel shows the amplitude of the LG velocity in successively larger rest frames as the velocity field of 2MRS is reconstructed at increasing radii. Incompleteness is illustrated by the black error bar on data points beyond  $120 h^{-1} \text{ Mpc}$  in the bottom panel. The top panel shows how the direction changes as the radius increases.

verified whether such a large convergence depth is compatible with a  $\Lambda$ CDM power-spectrum and a Gaussian initial condition (Juszkiewicz et al. 1990; Strauss et al. 1995).

G.L. and R.M. thank Jerry Ostriker for helpful suggestions and acknowledge travel grants from French ANR (OTARIE).

#### REFERENCES

- Cole, S., Percival, W. J., Peacock, J. et al. 2005, MNRAS  
 Colless, M. 1995, A.J., 109, 1937+  
 Crook, A. C., Huchra, J. P., Martimbeau, N. et al. 2007, ApJ, 655, 790  
 Dunkley, J., Komatsu, E., Nolte, M. R. et al. 2008, ArXiv e-prints, 803  
 Erdoğdu, P., Huchra, J. P., Lahav, O., et al. 2006a, MNRAS, 368, 51  
 Erdoğdu, P., Lahav, O., Huchra, J. P., et al. 2006b, MNRAS, 373, 45  
 Freedman, W. L., Madore, B. F., Gibson, B. K. et al. 2001, ApJ, 553, 47  
 Giovanelli, R., Haynes, M. P., Wegner, G. et al. 1996, ApJL, 464

- Huchra, J., Martimbeau, N., Jarrett, T., et al. 2005, in IAU Symposium, Vol. 216, Maps of the Cosmos, ed. M. Colless, L. Staveley-Smith, & R. A. Stathakis, 170–+
- Huchra, J. P. & Geller, M. J. 1982, ApJ, 257, 423
- Jackson, J. C. 1972, MNRAS, 156, 1P
- Juszkiewicz, R., Vittorio, N., & Wyse, R. F. G. 1990, ApJ, 349, 408
- Karachentsev, I. D., Makarov, D. I., Sharina et al. 2003, A&A, 398, 479
- Kocevski, D. D. & Ebeling, H. 2006, ApJ, 645, 1043
- Lauer, T. R. & Postman, M. 1994, ApJ, 425, 418
- Lavaux, G., Mohayaee, R., Colombi, S. et al. 2008, MNRAS, 383, 1292
- Lee, M. G., Freedman, W. L., & Madore, B. F. 1993, ApJ, 417, 553+
- Lin, Y. T., Mohr, J. J., & Stanford, S. A. 2004, ApJ, 610, 745
- Mohayaee, R., Mathis, H., Colombi, S. et al. 2006, MNRAS, 365, 939
- Mohayaee, R. & Tully, R. B. 2005, ApJL, 635, L113
- Muñoz, J. A. & Loeb, A. 2008, ArXiv e-prints, 805
- Ramella, M., Pisani, A., & Geller, M. J. 1997, AJ, 113, 483+
- Regos, E. & Szalay, A. S. 1989, ApJ, 345, 627
- Shaya, E. J., Peebles, P. J. E., & Tully, R. B. 1995, ApJ, 454, 15
- Strauss, M. A., Cen, R., Ostriker, J. P., Lauer, T. R., & Postman, M. 1995, ApJ, 444, 507
- Tegmark, M., Strauss, M. A., Blanton, M. R. et al. 2004, Phys. Rev. D, 69, 103501
- Tonry, J. & Schneider, D. P. 1988, AJ, 96, 807
- Tonry, J. L., Dressler, A., Blakeslee, J. P. et al. 2001, ApJ, 546, 681
- Tully, R. B. & Fisher, J. R. 1977, A&A, 54, 661
- Tully, R. B. & Pierce, M. J. 2000, ApJ, 533, 744
- Tully, R. B., Shaya, E. J., Karachentsev, I. D. et al. 2008, ApJ, 676, 184
- Zel'dovich, Y. B. 1970, A&A, 5, 84

# Bibliographie

- Bahcall, J. N., & Tremaine, S. 1981, *Astrophys. J.*, 244, 805
- Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, *ApJ*, 304, 15
- Bennett, C. L. et al. 2003, *ApJS*, 148, 1, arXiv :astro-ph/0302207
- Bernardeau, F. 1994, *ApJ*, 427, 51, astro-ph/9311066
- Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, *Phys. Rep.*, 367, 1, astro-ph/0112551
- Bertschinger, E., & Dekel, A. 1989, *ApJL*, 336, L5
- Bertsekas, D. 1998, *Network Optimization : Continuous and Discrete Models* (Athena Scientific)
- Bertsekas, D., & Tsitsiklis, J. 1997, *Parallel and Distributed Computation : Numerical Methods* (Athena Scientific)
- Bertsekas, D. P. 1979, *A Distributed Algorithm for the Assignment Problem* (MIT Press, Cambridge, MA)
- Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, *Astrophys. J.*, 379, 440
- Branchini, E., Eldar, A., & Nusser, A. 2002, *MNRAS*, 335, 53, astro-ph/0110618
- Brenier, Y. 1987, *C.R. Acad. Sci. Paris I*, 305, 805
- . 1991, *Comm. Pure Appl. Math.*, 44, 375
- Brenier, Y., Frisch, U., Hénon, M., Loeper, G., Matarrese, S., Mohayaee, R., & Sobolevskii, A. 2003, *MNRAS*, 346, 501, astro-ph/0304214
- Colless, M. et al. 2003, *ArXiv Astrophysics e-prints*, astro-ph/0306581
- Colombi, S., Chodorowski, M. J., & Teyssier, R. 2007, *MNRAS*, 375, 348
- Couchman, H. M. P., Thomas, P. A., & Pearce, F. R. 1995, *ApJ*, 452, 797, astro-ph/9409058
- Croft, R. A. C., & Gaztanaga, E. 1997, *Mon. Not. Roy. Astron. Soc.*, 285, 793, astro-ph/9602100
- Crook, A. C., Huchra, J. P., Martimbeau, N., Masters, K. L., Jarrett, T., & Macri, L. M. 2007, *ApJ*, 655, 790, arXiv :astro-ph/0610732
- Dekel, A., Bertschinger, E., & Faber, S. M. 1990, *ApJ*, 364, 349
- Dekel, A., Eldar, A., Kolatt, T., Yahil, A., Willick, J. A., Faber, S. M., Courteau, S., & Burstein, D. 1999, *Astrophys. J.*, 522, 1, astro-ph/9812197
- Diemand, J., Kuhlen, M., & Madau, P. 2007, *Astrophys. J.*, 657, 262
- Dunkley, J. et al. 2008, *ArXiv e-prints*, 803, 0803.0586
- Efstathiou, G., Bond, J. R., & White, S. D. M. 1992, *Mon. Not. Roy. Astron. Soc.*, 258, 1P



- Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1988, *MNRAS*, 235, 715
- Erdogdu, P. et al. 2006a, *ArXiv Astrophysics e-prints*, astro-ph/0605343
- Erdogdu, P. et al. 2006b, *MNRAS*, 373, 45, arXiv :astro-ph/0610005
- Ettori, S., Fabian, A. C., & White, D. A. 1997, *Mon. Not. Roy. Astron. Soc.*, 289, 787, astro-ph/9704092
- Falco, E. E. et al. 1999, *PASP*, 111, 438, astro-ph/9904265
- Fontanot, F., Monaco, P., & Borgani, S. 2003, *MNRAS*, 341, 692, astro-ph/0301601
- Freedman, W. L. 2000, *Phys. Rep.*, 333, 13, astro-ph/9909076
- Freedman, W. L. et al. 2001a, *Astrophys. J.*, 553, 47, astro-ph/0012376
- . 2001b, *Astrophys. J.*, 553, 47, astro-ph/0012376
- Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M., & Rossi, F. 2003, *Gnu Scientific Library : Reference Manual* (Network Theory Ltd.)
- Gurbatov, S. N., Saichev, A. I., & Shandarin, S. F. 1989, *MNRAS*, 236, 385
- Heisler, J., Tremaine, S., & Bahcall, J. N. 1985, *Astrophys. J.*, 298, 8
- Hénon, M. 2002, *ArXiv Mathematics e-prints*, math/0209047
- Huchra, J. et al. 2005, in *IAU Symposium*, Vol. 216, *Maps of the Cosmos*, ed. M. Colless, L. Staveley-Smith, & R. A. Stathakis, 170–+
- Huchra, J. P., & Geller, M. J. 1982, *ApJ*, 257, 423
- Huchra, J. P., Geller, M. J., Clemens, C. M., Tokarz, S. P., & Michel, A. 1992, *Bulletin d’Information du Centre de Donnees Stellaires*, 41, 31
- Huchra, J. P., Geller, M. J., de Lapparent, V., & Burg, R. 1988, in *IAU Symposium*, Vol. 130, *Large Scale Structures of the Universe*, ed. J. Audouze, M.-C. Pelletan, & S. Szalay, 105–+
- Jarrett, T. H., Chester, T., Cutri, R., Schneider, S., Skrutskie, M., & Huchra, J. P. 2000, *Astron. J.*, 119, 2498, astro-ph/0004318
- Jarrett, T. H., Chester, T., Cutri, R., Schneider, S. E., & Huchra, J. P. 2003, *Astron. J.*, 125, 525
- Johnson, H. L. 1966, *ARA&A*, 4, 193
- Jones, D. H. et al. 2004, *Mon. Not. Roy. Astron. Soc.*, 355, 747, astro-ph/0403501
- Jones, D. H., Saunders, W., Read, M., & Colless, M. 2005, *Publications of the Astronomical Society of Australia*, 22, 277, astro-ph/0505068
- Juszkiewicz, R., Vittorio, N., & Wyse, R. F. G. 1990, *Astrophys. J.*, 349, 408
- Kaiser, N. 1987, *MNRAS*, 227, 1
- Kaiser, N. 1988, *Mon. Not. Roy. Astron. Soc.*, 231, 149
- Karachentsev, I. D. et al. 2006, *Astron. J.*, 131, 1361, astro-ph/0511648
- Karachentsev, I. D., Karachentseva, V. E., Huchtmeier, W. K., & Makarov, D. I. 2004, *Astron. J.*, 127, 2031

- Karachentsev, I. D. et al. 2003, *Astron. Astrophys.*, 398, 479, astro-ph/0211011
- . 2002, *Astron. Astrophys.*, 389, 812, astro-ph/0204507
- Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, *Mon. Not. Roy. Astron. Soc.*, 303, 188, astro-ph/9805283
- Klypin, A. 2000, *ArXiv Astrophysics e-prints*, astro-ph/0005502
- Kocevski, D. D., & Ebeling, H. 2006, *Astrophys. J.*, 645, 1043, astro-ph/0510106
- Lahav, O., Fisher, K. B., Hoffman, Y., Scharf, C. A., & Zaroubi, S. 1994, *ApJL*, 423, L93+, arXiv :astro-ph/9311059
- Lee, M. G., Freedman, W. L., & Madore, B. F. 1993, *Astrophys. J.*, 417, 553+
- Lin, Y. T., Mohr, J. J., & Stanford, S. A. 2004, *Astrophys. J.*, 610, 745, astro-ph/0402308
- Marinoni, C., & Hudson, M. J. 2002, *ApJ*, 569, 101, astro-ph/0109134
- Mohayaee, R., Mathis, H., Colombi, S., & Silk, J. 2006, *MNRAS*, 365, 939, astro-ph/0501217
- Mohayaee, R., & Tully, R. B. 2005, *ApJL*, 635, L113, astro-ph/0509313
- Monaghan, J. J. 1992, *Annual review of astronomy and astrophysics*, 30, 543
- Nusser, A., & Branchini, E. 2000, *Mon. Not. Roy. Astron. Soc.*, 313, 587, astro-ph/9908167
- Nusser, A., & Colberg, J. M. 1998, *Mon. Not. Roy. Astron. Soc.*, 294, 457+, astro-ph/9705121
- Peebles, P. J. E. 1980, *The large-scale structure of the universe* (Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.)
- . 1989, *ApJL*, 344, L53
- Phelps, S. D., Desjacques, V., Nusser, A., & Shaya, E. J. 2006, *MNRAS*, 370, 1361, arXiv :astro-ph/0512405
- Pizagno, J. et al. 2006, *ArXiv Astrophysics e-prints*, astro-ph/0608472
- Press, W. H., & Schechter, P. 1974, *ApJ*, 187, 425
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical recipes in C. The art of scientific computing* (Cambridge : University Press, —c1992, 2nd ed.)
- Ramella, M., Pisani, A., & Geller, M. J. 1997, *Astron. J.*, 113, 483+
- Regos, E., & Szalay, A. S. 1989, *Astrophys. J.*, 345, 627
- Sarkar, Devdeep, Feldman, Hume, A., Watkins, & Richard. 2007, *Monthly Notices of the Royal Astronomical Society*, 375, 691, astro-ph/0607426
- Shandarin, S. F., & Zeldovich, Y. B. 1989, *Reviews of Modern Physics*, 61, 185
- Shaya, E. J., Peebles, P. J. E., & Tully, R. B. 1995, *ApJ*, 454, 15, astro-ph/9506144
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *Mon. Not. Roy. Astron. Soc.*, 323, 1, astro-ph/9907024
- Sheth, R. K., & Tormen, G. 2002, *MNRAS*, 329, 61, astro-ph/0105113
- Spergel, D. N. et al. 2006, *ArXiv Astrophysics e-prints*, astro-ph/0603449

- Stoughton, C. et al. 2002, *Astron. J.*, 123, 485
- Strauss, M. A., & Willick, J. A. 1995, *Phys. Rept.*, 261, 271, astro-ph/9502079
- Tegmark, M. et al. 2004a, *ApJ*, 606, 702, arXiv :astro-ph/0310725
- . 2006, *Phys. Rev. D*, 74, 123507, arXiv :astro-ph/0608632
- . 2004b, *Phys. Rev. D*, 69, 103501, arXiv :astro-ph/0310723
- Thorne, K. S., Misner, C. W., Wheeler, J. A., Thorne, K., & Wheeler, J. 1973, *Gravitation (Physics Series)* (W. H. Freeman)
- Tonry, J., & Schneider, D. P. 1988, *Astron. J.*, 96, 807
- Tonry, J. L., Dressler, A., Blakeslee, J. P., Ajhar, E. A., Fletcher, A. B., Luppino, G. A., Metzger, M. R., & Moore, C. B. 2001, *Astrophys. J.*, 546, 681, astro-ph/0011223
- Tully, R. B. 1987, *ApJ*, 321, 280
- . 2005, *ApJ*, 618, 214, astro-ph/0312441
- Tully, R. B., & Fisher, J. R. 1977, *Astromy and Astrophysics*, 54, 661
- Tully, R. B., & Pierce, M. J. 2000, *ApJ*, 533, 744
- Tully, R. B., & Pierce, M. J. 2000, *Astrophys. J.*, 533, 744
- Tully, R. B., Shaya, E. J., Karachentsev, I. D., Courtois, H., Kocevski, D. D., Rizzi, L., & Peel, A. 2008, *Astrophys. J.*, 676, 184, 0705.4139
- Uzan, J.-P., & Peter, P. 2005, *Cosmologie primordiale (Belin-2 Scolaire & Universitai)*
- Valentine, H., Saunders, W., & Taylor, A. 2000, *MNRAS*, 319, L13
- Wald, R. M. 1984, *General Relativity* (University Of Chicago Press)
- Zandivarez, A., Martínez, H. J., & Merchán, M. E. 2006, *ApJ*, 650, 137, astro-ph/0602405
- Zaroubi, S., Zehavi, I., Dekel, A., Hoffman, Y., & Kolatt, T. 1997, *Astrophys. J.*, 486, 21+, astro-ph/9610226
- Zel'Dovich, Y. B. 1970, *A&A*, 5, 84

# Table des figures

2.1	Catalogue LICK . . . . .	4
2.2	Catalogue CfA2 . . . . .	5
2.3	Cosmographie locale . . . . .	7
2.4	Illustration des régions du spectre de puissance des fluctuations qui sont observées par différentes expériences . . . . .	19
2.5	Mesure de la constante de Hubble grâce aux céphéides . . . . .	20
2.6	Diagramme illustrant le processus général de reconstruction du champ de vitesse . . .	21
2.7	Illustration de la méthode de reconstruction des champs de vitesse POTENT . . . . .	23
2.8	Exemple d'orbites reconstruites par la méthode Fast-Action-Method . . . . .	25
2.9	La méthode de reconstruction MAK . . . . .	27
3.1	Champ de densité d'une simulation . . . . .	30
3.2	Fonction de masse de la simulation . . . . .	32
3.3	Illustration de la reconstruction d'une composante du champ de déplacement . . . . .	33
3.4	Comparaison champ déplacement reconstruit/simulé . . . . .	34
3.5	Construction d'un maillage de MAK . . . . .	35
3.6	Champ de déplacement de FullMock . . . . .	36
3.7	Volume lagrangien projeté selon l'axe Z d'un groupe . . . . .	36
3.8	Reconstruction du champ de vitesse de la simulation . . . . .	38
3.9	Reconstructin du champ de vitesse sur FullMock . . . . .	39
3.10	Distribution cumulée d'erreur de reconstruction de la direction du champ de vitesse .	40
3.11	Tentative de reconstruction des conditions initiales avec la théorie linéaire . . . . .	41
3.12	Conditions initiales reconstruites . . . . .	43
3.13	Comparaison condition initiales originales/reconstruites . . . . .	44
3.14	Illustration de l'effondrement sphérique . . . . .	45
3.15	Effondrement sphérique et conditions initiales . . . . .	46
3.16	Reconstruction de la densité de probabilité de fluctuations sur la simulation . . . . .	48
3.17	Reconstruction de la densité de probabilité de fluctuations sur FullMock . . . . .	49
3.18	Comparaison de différentes correction de la PDF . . . . .	49
4.1	Statistique des erreurs des les déplacements reconstruits . . . . .	53
4.2	Statistique des erreurs des vitesses reconstruites . . . . .	54
4.3	Analyse en maximum de la fonction de vraisemblance . . . . .	57
4.4	Biais statistique . . . . .	63
5.1	Evaluation de la masse diffuse . . . . .	70
5.2	Correction de l'effet de masse diffuse . . . . .	71
5.3	Choix du rapport $M/L$ . . . . .	72
5.4	Fonction $M/L$ . . . . .	74
5.5	Biais de $M/L$ . . . . .	75
5.6	Limitation en magnitude/Correction de la masse manquante . . . . .	79
5.7	Limitation en magnitude/Fraction de luminosité manquante . . . . .	80
5.8	Incomplétude : limitation en magnitude, résultat . . . . .	81
5.9	Correction de l'effet de distortion de décalage en rouge . . . . .	84
5.10	Distribution des erreurs du champ de vitesse reconstruit, espace redshift . . . . .	85

5.11	Illustration de l'algorithme de remplissage du plan galactique . . . . .	87
5.12	Simulation du plan galactique de la voie lactée . . . . .	88
5.13	Obscurcissement du plan galactique / champ de vitesse . . . . .	89
5.14	Domaine lagrangien / sans distortions de redshift . . . . .	90
5.15	Domaine lagrangien / avec distortions de redshift . . . . .	91
5.16	Volume lagrangien d'un catalogue . . . . .	92
5.17	Variance cosmique . . . . .	94
5.18	Evolution de $\Omega_{m,mes}$ en fonction de l'environnement . . . . .	96
5.19	Empilement des effets observationnels . . . . .	99
6.1	Distribution de masse dans le catalogue NBG-3k . . . . .	105
6.2	Biais dans la construction du maillage de MAK pour NBG-3k . . . . .	106
6.3	Flot des vitesses reconstruites pour le catalogue NBG-3k . . . . .	107
6.4	Projection aitoft du 2MASS Redshift survey . . . . .	108
6.5	Distribution en redshift des galaxies du 2MRS . . . . .	109
6.6	Distribution en luminosité des galaxies du 2MRS . . . . .	111
6.7	Comptage cumulé en redshift des galaxies du 2MRS . . . . .	112
6.8	Projection aitoft des galaxies du 2MASS Redshift survey ayant une mesure de distance . . . . .	113
6.9	Compression des finger-of-god dans le 2MRS . . . . .	113
6.10	Représentation imagée des paramètres de l'algorithme de compression des finger-of-god . . . . .	115
6.11	Projection aitoft et cartésienne du 2MRS après compression des finger-of-god . . . . .	116
6.12	Galaxies du volume local dans le 2MRS . . . . .	117
6.13	Comparaison magnitude totale apparente obtenue par extrapolation à la magnitude $K_{20}$ . . . . .	118
6.14	Estimation du rapport $M/L$ des groupes de galaxies du 2MRS . . . . .	119
6.15	Remplissage du plan galactique du 2MRS . . . . .	120
6.16	Représentant du catalogue 2MRS après remplissage du plan galactique avec l'algorithme de la section 5.4.1. . . . .	121
6.17	Correction de l'incomplétude dans le catalogue 2MRS . . . . .	122
6.18	Correction de l'incompétude : fonction d'amplification de la luminosité . . . . .	123
6.19	Comparaison du champ de densité du NBG-3k à celui du 2MRS . . . . .	124
6.20	Vitesses propres reconstruites du 2MRS (reconstruction $120 h^{-1}\text{Mpc}$ ) . . . . .	126
6.21	Vitesses propres reconstruites du 2MRS, position réelle . . . . .	127
6.22	Champ de vitesse lissée adaptativement sur grille . . . . .	128
6.23	Test de la procédure sur une simulation . . . . .	131
6.24	Estimation de bulk-flow dans une simulation . . . . .	132
6.25	Origine du mouvement du Groupe Local par rapport au CMB . . . . .	138
6.26	Illustration du champ de densité à différentes distances de l'observateur . . . . .	139
6.27	Origine du mouvement du superamas local (volume de rayon $30 h^{-1}\text{Mpc}$ ) . . . . .	143
6.28	Origine du mouvement du Groupe Local par rapport au super-amas local . . . . .	144
6.29	Comparaison vitesses reconstruites/simulées dans le 2MASS-Mock . . . . .	146
6.30	Illustration de l'effet du bulk flow sur la comparaison vitesse-vitesse . . . . .	146
6.31	Comparaison des vitesses reconstruites avec 2MASS aux vitesses observées . . . . .	148
B.1	Machine de Hénon . . . . .	158
B.2	Complexité temporelle de l'algorithme de "Auction" . . . . .	161
C.1	Transformées de Fourier de fonctions de filtrages usuelles . . . . .	166
C.2	Dispersion du champ de vitesse . . . . .	167
C.3	Variance entre bulk flow estimé par la méthode $\chi^2$ et théorique . . . . .	169
D.1	Fonction de masse de Press & Shechter et de Sheth & Tormen . . . . .	172
E.1	Nombre de particules par sites de grille . . . . .	177

E.2	Illustration du lissage adaptatif pour l'obtention d'un champ de densité d'une simulation	178
-----	---	-----



# Liste des tableaux

2.1	Cosmographie locale . . . . .	6
2.2	Paramètres cosmologiques donnés WMAP5 . . . . .	13
5.1	Effet de biais de M/L . . . . .	74
5.2	Incomplétude : limitation en magnitude . . . . .	80
5.3	Reconstruction en redshift . . . . .	80
5.4	Effet du plan galactique dans la reconstruction sur catalogues virtuels . . . . .	87
5.5	Volume lagrangien . . . . .	95
5.6	Variance cosmique . . . . .	97
6.1	Catalogue virtuel / Mesure de l'effet "dipole" du CMB . . . . .	133
6.2	Catalogue virtuel / Mesure et reconstruction de bulk flow $3000 \text{ km s}^{-1}$ dans un catalogue virtuel . . . . .	134
6.3	Reconstruction du mouvement du Groupe Local . . . . .	134
6.4	Mesure et reconstruction de bulk flow $3000 \text{ km s}^{-1}$ à l'aide du catalogue NBG-3k . . . . .	135
6.5	Origine du mouvement du Groupe Local par rapport au CMB . . . . .	137
6.6	Origine du mouvement du superamas local (volume de rayon $30 h^{-1} \text{Mpc}$ ) . . . . .	142
C.1	Meilleur ajustement analytique du spectre de puissance . . . . .	164
C.2	Quelques fonctions de filtrage usuelles . . . . .	165